







# MONOGRAPHS ON PHYSICS

EDITED BY

SIR J. J. THOMSON, O.M., F.R.S.

MASTER OF TRINITY COLLEGE, CAMBRIDGE

AND

FRANK HORTON, Sc.D.

PROFESSOR OF PHYSICS IN THE UNIVERSITY OF LONDON



## MONOGRAPHS ON PHYSICS.

Edited by SIR J. J. THOMSON, O.M., F.R.S.,  
Master of Trinity College, Cambridge,  
and FRANK HORTON, Sc.D.,  
Professor of Physics in the University of London.

THE SPECTROSCOPY OF THE EXTREME ULTRA-VIOLET. By THEODORE LYMAN, Ph.D., Assistant Professor of Physics in Harvard University. With Diagrams.

RELATIVITY, THE ELECTRON THEORY, AND GRAVITATION. By E. CUNNINGHAM, M.A., Fellow and Lecturer of St. John's College, Cambridge.

THE EMISSION OF ELECTRICITY FROM HOT BODIES. By O. W. RICHARDSON, F.R.S., Wheatstone Professor of Physics, King's College, London.

MODERN SEISMOLOGY. By G. W. WALKER, A.R.C.Sc., F.R.S., Deputy University Lecturer in Astrophysics in the University of Cambridge. With Plates and Diagrams.

RAYS OF POSITIVE ELECTRICITY AND THEIR APPLICATION TO CHEMICAL ANALYSIS. By SIR J. J. THOMSON, O.M., F.R.S., Master of Trinity College, Cambridge. With Illustrations.

---

LONGMANS, GREEN AND CO.

39 PATERNOSTER ROW, LONDON

NEW YORK, BOMBAY, CALCUTTA, AND MADRAS

RELATIVITY  
THE ELECTRON THEORY  
AND  
GRAVITATION

BY  
E. CUNNINGHAM, M.A.  
FELLOW AND LECTURER OF ST. JOHN'S COLLEGE, CAMBRIDGE

*WITH DIAGRAMS*

*SECOND EDITION*

LONGMANS, GREEN AND CO.  
39 PATERNOSTER ROW, LONDON  
FOURTH AVENUE & 30TH STREET, NEW YORK  
BOMBAY, CALCUTTA, AND MADRAS

1921



## PREFACE TO SECOND EDITION.

THE first edition of this book was published while the General Principle of Relativity was being worked out, before it seemed possible to arrive at any confirmation from observation. Shortly after, however, it was shown that the new theory explained the motion of the perihelion of Mercury, and now the result of the Solar Eclipse expedition has clinched matters.

It seemed best to leave practically untouched the account of the special principle as the natural avenue of approach to the broader view, which is described at length in Part II.

Acknowledgment is due to the author's friends, Prof. H. F. Baker and Mr. E. V. Appleton, who kindly read through the proofs of the new matter.

E. C.

CAMBRIDGE, 1920.

## PREFACE TO FIRST EDITION.

THIS monograph is an attempt to set out as clearly and simply as possible the relation of the Principle of Relativity to the generally accepted Electron Theory, showing at what points the former is the natural and necessary complement of the latter.

No attempt has been made to describe to any great extent consequences of the Principle which would be for the most part beyond the reach of experimental investigation, and the mathematical analysis has been omitted as far as possible with the hope of rendering the account useful to the general reader, especially to the experimental physicist. Those who desire to follow out the train of thought in more detail, and especially to make acquaintance with the mathematical presentation developed by Minkowski, may be referred to the author's larger book on the same subject. The author's acknowledgments are due to the Cambridge University Press for their permission to use some of the material of that work in the preparation of this.

E. C.

CAMBRIDGE, 1915.

# CONTENTS.

	PAGE
Preface to Second Edition . . . . .	v
Preface to First Edition . . . . .	vi
Note on Vector Notation . . . . .	i
<b>CHAPTER</b>	
I. Introductory . . . . .	2

## PART I.

### THE SPECIAL PRINCIPLE.

II. The Origin of the Principle . . . . .	8
III. The Relativity of Space and Time . . . . .	28
IV. The Relativity of the Electro-Magnetic Vectors . . . . .	46
V. Mechanics and the Principle of Relativity . . . . .	61
VI. Minkowski's Four-Dimension Vectors . . . . .	72

## PART II.

### THE GENERAL PRINCIPLE OF RELATIVITY.

VII. The General Theory . . . . .	86
VIII. Verification of Einstein's Theory . . . . .	113
IX. Further Generalization, Weyl's Theory of Electricity . . . . .	125
Index . . . . .	147



## NOTES ON VECTOR NOTATION.

VECTOR quantities are denoted throughout the book by Clarendon Type, thus, **e**, **h**, **u**.

The components of vectors are denoted by italic type, thus **e** has components ( $e_x, e_y, e_z$ ), **F** has components ( $F_x, F_y, F_z$ ).

### Definitions.

(i) *Vector Product*.—The vector product of two vectors, **e**, **h**, is a vector at right angles to the directions of both, and of magnitude equal to the product of the magnitudes of the two into the sine of the angle between them. It is denoted by the symbol  $[\mathbf{eh}]$ .

The components of  $[\mathbf{eh}]$  are

$$(e_y h_z - e_z h_y, e_z h_x - e_x h_z, e_x h_y - e_y h_x).$$

(ii) *Divergence* of a vector.

$$\begin{aligned} \operatorname{div} \mathbf{e} &= \frac{\partial e_x}{\partial x} + \frac{\partial e_y}{\partial y} + \frac{\partial e_z}{\partial z} \\ &= \lim \iint e_n dS / V \text{ as } V \rightarrow 0 \end{aligned}$$

where  $dS$  is an element of a closed surface bounding a volume  $V$ , and  $e_n$  is the component of **e** normal to  $dS$ .

(iii) *Curl* of a vector.

$$\operatorname{curl} \mathbf{e} = \left( \frac{\partial e_z}{\partial y} - \frac{\partial e_y}{\partial z}, \frac{\partial e_x}{\partial z} - \frac{\partial e_z}{\partial x}, \frac{\partial e_y}{\partial x} - \frac{\partial e_x}{\partial y} \right).$$

The component of  $\operatorname{curl} \mathbf{e}$  in a given direction is

$$\text{equal to } \lim \oint e_s ds / S \text{ as } S \rightarrow 0$$

where  $ds$  is an element of the arc of a closed curve which bounds a small area  $S$  to which the given direction is normal.



## CHAPTER I.

### INTRODUCTORY.

1. THE term Principle of Relativity was first applied to the hypothesis that it is impossible by means of physical experiments to determine the absolute velocity of a body through space. This was introduced by Einstein in 1905 not as a metaphysical doctrine based on the supposition that it is impossible to conceive of an absolute standard of position in space. Such an impossibility had often been asserted by writers on the theory of mechanics, but it was quite irrelevant to Einstein's hypothesis. Physical theory has nothing to do with an absolute space. The space with which it deals is one aspect of the relations which have been observed between the phenomena with which that theory is concerned.

Newton in setting out his scheme of dynamics postulated an absolute standard of position in space relative to which all velocities are measured; such a beginning was suggested by the history of the astronomical problem with which he was dealing. Copernicus shifted the centre of the universe from the earth to the sun, and from this it was a very short step to the assumption that the sun was no more likely to be an absolute centre than the earth. Hence the assumption of an unknown, though still definite, criterion of position in space.

But when, on the background of this assumption, the science of dynamics had been developed, it was found that the laws which had been framed had a very special property. If they were universally satisfied, there was no means of determining the actual velocity of any body, though the relative velocity of two bodies was determinate. In fact, it in no way disturbs the

general laws of dynamics if an arbitrary velocity is added to the velocity of every body in the system. In other words, the frame of reference assumed at the beginning is not a unique one, but may be any one of an infinite number, of which any one has relative to any other a constant velocity of translation without rotation. To assert that a unique frame of reference can never be found is perhaps not absolutely proved, but neither is any scientific hypothesis. All that has been done is that the hypothesis has been put on such a footing that it seems a waste of time to endeavour to dispense with it.

This dynamical relativity is a very different thing from a statement as to what the mind is able to conceive about position and motion. It is indeed doubtful whether it is possible or logical to think about either of these entirely apart from the regularities perceived in the motions of actual bodies. In any case, if the doctrine of the relativity of all motion were *meta*-physical, it would have to extend beyond the scope of that indicated by Newton's laws of motion, and to deny the existence of any criterion of a fixed direction in space or of the validity of the conception of a constant velocity. It would result in an entirely agnostic position about motion which would have to be superseded whenever it was desired to deal with actual facts of experience. We shall see later that Einstein has in his later work been able to show that the facts, as they are known to us, are as a matter of fact entirely consistent with such a position.

2. The limited scope of the relativity arising out of dynamical theory has always been rather unsatisfying to the mind, and so it was almost with a sense of relief that the rise of the electro-magnetic theory of light, bringing with it the suggestion of a universal æther, was hailed as offering a possibility of a return to an absolute theory, that is, to one in which a velocity can be uniquely assigned to every body, namely, its velocity relative to the æther. In fact, it seemed clear, since all astronomical observations are made by optical means, that the frame of reference in practice must be actually the medium relative to which light is propagated. Thus arose the attempt to determine the velocity of the earth relative to the æther.

Such a quest has of course no meaning apart from a definite conception of the æther. In the first instance the experiments were rather directed towards clearing up the question as to what was the best way to think of the æther, whether as another variety of matter filling all space not otherwise occupied, or as a medium of such nature that it could penetrate matter and be undisturbed by the passage of matter through it. It was the latter way of thinking of it toward which opinion gradually tended and which had been almost universally adopted. It is perhaps not wise to make a definite statement as to why this was so, but, at any rate, a very important factor was the development of the precise statement of the laws of electro-magnetic phenomena as initiated by Maxwell and elaborated in the electron theory of Larmor and Lorentz.

Just as the enunciation of Newton's laws of motion gave a meaning to the notion of 'absolute direction' which before it did not possess, so the adoption of Maxwell's equations, as expressing the laws of propagation of light and electro-magnetic disturbances, gave a new meaning to the term 'position in space' by relating it to the medium whose properties the equations were thought to embody. We must remember, however, that this medium is defined only by these same properties and equations, and that to base any argument on an analogy between it and a hypothetical continuous medium of material and mechanical properties in the ordinary sense—properties which have not been shown to be involved in the definition of the æther—is to become liable to finding ourselves at variance with the facts. The æther is, in truth, nothing more than the aggregate of the functions which it serves, though we find it difficult to think of it except in some concrete and uniquely existent form; the exact form, however, which is adopted is largely determined by individual preference for this or that analogy.

3. The conception of the 'stagnant' æther having been adopted, it came as a great surprise that all experimental efforts failed to find a physical effect due to the motion of bodies through it. Theory seemed to indicate several ways of observing such an effect. The failure of all these methods

necessitated a reconciliation of theory and experiment, and such a reconciliation was rendered possible by the work of Larmor and Lorentz ; if not complete in every detail, yet it went so far that many were led to the belief that not only some but all experiments of this nature were foredoomed to failure (e.g. Larmor, Brit. Ass., Belfast, 1902).

To admit such a possibility is to allow that the æther, whatever its nature, may for ever remain concealed. This is to weaken considerably our power of thinking of the æther as an actually existing medium. But it in no way affects the scheme of properties which it was intended to connote. Rather in the hands of Lorentz and Larmor it becomes a confirmation of the theory to find that the experiments actually do fail. But it is hardly true to say that, if we grant the hypothesis that we shall never be able to identify a unique frame of reference for the mathematical theory, then we deny the existence of a real medium of propagation of the physical effects of light and electricity. What we do is only to question the sufficiency of the existing conceptions of the nature of this medium, and the validity of identifying it with a frame of reference which experience finds to be far from definite.

4. If the dynamics of Newton left some ambiguity in the meaning of space relations there is no such doubt left in the case of time intervals. To him a time interval was so definite that he was able to speak of absolute time as 'flowing evenly on' as if independent of all phenomena. One of the fundamental elements in this idea of absolute time is that of the 'simultaneity of events' at different places. No ambiguity about the meaning of this appears to have arisen.

But if we grant the impossibility of determining the velocity of the earth through the æther, using the term for a moment in the sense of the frame of reference for the propagation of light, this idea of 'simultaneity' becomes as vague as that of the velocity of a moving body. If we consider that the science of astronomy and the law of gravitation are dependent for their verification on optical observations, we shall be inclined to agree that so far at any rate the criterion of simultaneity which has been used in practice has been one

based on optical communication. Now we shall see later that the setting up of a standard of simultaneity by means of light signals is not possible until a definite velocity is assigned to the observer. Thus the hypothesis of relativity requires a reconsideration of the way in which we measure time.

This again reacts on the measurement of the length of a material body, the 'distance between two points' being the distance between simultaneous positions of those points. Thus it becomes necessary also to re-examine the way in which we measure space. It becomes impossible to consider space and time separately; the two measures are inter-related to such an extent that Minkowski felt himself constrained to say that "from henceforth time by itself and space by itself are mere shadows, that they are only two aspects of a single and indivisible manner of co-ordinating the facts of the physical world".<sup>1</sup>

5. In what follows, then, we have to consider first the exact way in which our fundamental notions of space, time, and motion have to be modified. This will lead us on to a statement of the status assigned to dynamical theory according to the principle of relativity, and in particular will require a discussion of the ideas of momentum and energy. It is well known that the idea of a constant mass for every body has to be modified if we take into account the effects of radiation and electrical constitution. We shall see that we can draw, from the general Principle of Relativity, some very important conclusions as to the mechanical relations of systems, without having recourse to any particular theory of the nature of an electron or of the way in which matter is built up out of them. Our growing sense of the insufficiency of the existing pictures of the constitution of matter makes this a very important consideration.

The following out in detail of the consequences of Einstein's work of 1905, should by loosening our hold on the Newtonian methods of thought, so deeply ingrained, make it easier to take the further step to the appreciation of the much more general and satisfying position which underlies Einstein's more recent

<sup>1</sup> *Raum und Zeit*, Leipzig, 1909.

work. To this we shall return later. For the moment, we can but record the fact, that, setting out to examine the objection of the philosopher to the extremely limited scope of his so-called Principle of Relativity, he not only was able to satisfy the philosopher, but at the same time to explain facts which had refused to bow to Newtonian dynamics, and to lead to the entirely new discovery that rays of light are bent by gravitation. To this we shall come in Part II.

PART I.

THE SPECIAL PRINCIPLE OF RELATIVITY.

CHAPTER II.

THE ORIGIN OF THE PRINCIPLE.

6. SPACE AND TIME IN NEWTON'S DYNAMICS.

IN order to appreciate the full significance of the new point of view introduced by Einstein in 1905 with the technical name, "The Principle of Relativity" now more precisely called "*The Special Principle of Relativity*," it is desirable to make a survey of the previous lines of thought.

Newton planted deeply in scientific thought the ideas of an absolute position in space and of an absolute time. The significance of the Copernican revolution in astronomy extends beyond the moving of the origin of reference from the observer to the sun. It makes it impossible to think even of the sun as a permanent and ultimate point of reference. We naturally go on to think of the sun as one member of a system greater than the solar system, and we can see no end to the process of shifting the point of reference. We have no conception of the actual velocity of the earth or of the sun in space. But the precise scheme of laws set forth by Newton introduced an absolute element into our conceptions of celestial motions. In the light of his laws we perceive that the reason that we are able over a large range of terrestrial phenomena to treat the earth as the point of reference and to ignore the remainder of the universe, is just that we may think of the earth's absolute velocity as changing very slowly. Similarly, the reason that a theory of the solar system as a separate system is possible, ignoring

the whole of the rest of the universe, is that we may conceive of its centre of mass as having a constant absolute velocity.

Although later we find that the 'absolute velocity' of a body is not determinable, yet the classical dynamics do give us a criterion of something which we may call 'absolute direction,' though the philosopher may assert that we are not using the term in a strict sense. In the sense in which we use the term this absolute direction is not a thing that is obvious *a priori*, arising out of pure thought; it is a fact which emerges out of the perceived motions of bodies as we try to reduce them to the simplest possible descriptive scheme. The scientific conception of space is inseparable from the laws of physics. There is no other way of making space an object of measurement save by means of this body of law. Apart from it or from some equivalent, motion is a mere sensation, though it is doubtful whether the most primitive man is without some rudimentary notion of those regularities which have been summed up in the law of inertia.

#### 7. THE RELATIVITY OF NEWTON'S DYNAMICS.

But now we come to the remarkable fact that the absolute-ness introduced into our ideas of space by the laws of dynamics is only partial. It is a commonplace that if any frame of reference is known with respect to which the laws of motion are satisfied then any other frame which is moving relative to that known frame with a constant velocity as judged by the standards of dynamics, is equally suitable as a basis for the description of the motion by the same laws.

This partial relativity it will be convenient to speak of as 'Newtonian Relativity'. It is not a philosophic principle, it is purely empirical, and rests for its justification on the agreement of the deductions from it with the facts of observation. Thus from this point of view we have no criterion whereby we may say in any defined sense that a given point is at rest in space.

A similar position is to be taken as to the nature of time as an object of measurement. There may or may not be an *a priori* standard of the equality of two elements of duration, but we do not know what it may be. For scientific purposes



of exact measurement the time of which we think is defined by the laws of motion. There is no real need to say with Newton that "absolute time flows uniformly on," and to pretend that, given this time, we lay our phenomena of motion in thought against it and so measure them. We cannot doubt that the definition of time was invented after the system of dynamics had been worked out, and was prefixed to satisfy the demand for definitions and postulates at the beginning of any systematic treatise.

#### 8. TIME NOT AN ABSOLUTE OR INDEPENDENT CONCEPT.

Time, then, as an object of measurement is nothing more than one aspect of the relations which we have been able to disentangle from the complex of phenomena of motion. But to our everyday life it is such a definite aspect that it seems to exist almost in its own right. The partial relativity that exists in the scheme of relations leaves no ambiguity in the term 'equality of two intervals of time' nor in the term 'equality of two distances'. These are such definite and separable elements in the construction of the scheme that it is possible to reconstruct the idea of the measurement of motion out of them. The Newtonian picture of the motions of bodies is the perfected result of a long process of analysis of perceptions and refinement of concepts. Beginning with the crude perception of change, the continual comparison of motions results in the extraction of the conceptual time and space of mathematics. In terms of these we are able to build up a mental model of the motion which is susceptible to mathematical treatment. This model is only approximate, since it necessarily ignores certain factors in the actual system. Our planetary theory, for instance, disregards the effect of the distant stars, and the neglect is justified *a posteriori* by the fact that the discrepancy between model and system is incapable of observation.

#### 9. MATHEMATICAL FORM OF THE NEWTONIAN PRINCIPLE OF RELATIVITY.

Before proceeding further, it will be useful to set down in mathematical form the chief aspects of 'Newtonian Relativity'.

(i) *The Space-Time Transformations.*—Let  $(x, y, z)$  be a set of space co-ordinates and  $t$  an associated time variable, defining together a Newtonian frame of reference—that is, which are such that if the motion is described by the use of these variables, the laws of motion are satisfied. Let a new set of variables be taken, defined by the equations

$$x' = x - at, \quad y' = y - \beta t, \quad z' = z - \gamma t, \quad t' = t.$$

Now let the motion be described by turning the given relations between  $(x, y, z)$  and  $t$  into relations between  $(x', y', z')$  and  $t'$ , then the relations so obtained between  $(x', y', z')$  and  $t'$  will also satisfy the laws of motion; in other words, the new variables also define a Newtonian frame of reference.

A point which as measured in the first set of co-ordinates has the velocity  $(u, v, w)$  has in the second the velocity  $(u - a, v - \beta, w - \gamma)$ , and the relative velocity of two points is the same in both systems,  $a, \beta, \gamma$  being given constants.

(ii) *The Law of Addition of Velocities.*—The relations

$$u' = u - a, \quad v' = v - \beta, \quad w' = w - \gamma$$

connecting the velocities of a given point relative to two Newtonian frames of reference, of which the origin of the second has a velocity  $(a, \beta, \gamma)$  relative to the first, seem to us to be obvious relations; but it is necessary to remember that the fact that either  $(u, v, w)$  or  $(u', v', w')$  may be conveniently called the velocity of the body in the appropriate frame of reference depends upon what has been said of the possibility of maintaining unchanged the form of the dynamical laws when we change the frame of reference. We shall see later that these relations cease to be convenient when we are dealing with the extremely great velocities which we associate with the negative electrons constituting the cathode and  $\beta$ -rays.

(iii) Newtonian Relativity consists in the fact that either  $(x, y, z, t)$  or  $(x', y', z', t')$  are equally valid as space-time co-ordinates, and  $(u, v, w)$  or  $(u', v', w')$  as the corresponding measures of velocity. But there are certain quantities which are 'invariants,' that is, which have the same value whichever set of co-ordinates we employ. These are—to mention the most simple—the 'mass' of a particle, the 'force' acting upon

a particle, and the 'acceleration' of a particle. When we come to higher dynamics, we find that the 'constants of inertia' of a rigid body and other quantities have invariant values, whereas the 'energy' of a system is a *relative* quantity though the 'principle of the conservation of energy' is an *invariant relation*.

#### 10. THE PRINCIPLE OF CONSERVATION OF MOMENTUM DEDUCIBLE FROM THE PRINCIPLES OF CONSERVATION OF ENERGY AND OF RELATIVITY.

If we assume that the principle of the conservation of energy in the form

*'rate of increase of kinetic energy = rate of work of forces'*

is true whatever Newtonian frame of reference we use, we may deduce the principle that

*'rate of increase of momentum in any direction = sum of forces in that direction'*.

For, whatever the values of the constants  $\alpha, \beta, \gamma$ , we must have

$$\begin{aligned} \frac{d}{dt} \sum \frac{1}{2} m \{ (u - \alpha)^2 + (v - \beta)^2 + (w - \gamma)^2 \} \\ = \sum \{ X(u - \alpha) + Y(v - \beta) + Z(w - \gamma) \}, \end{aligned}$$

$m$  being the mass of a particle of the system, and  $(X, Y, Z)$  being the components of any force applied to the system.

Hence

$$\begin{aligned} \sum m \left\{ (u - \alpha) \frac{du}{dt} + (v - \beta) \frac{dv}{dt} + (w - \gamma) \frac{dw}{dt} \right\} \\ = \sum \{ X(u - \alpha) + Y(v - \beta) + Z(w - \gamma) \}, \end{aligned}$$

whatever the values of  $\alpha, \beta, \gamma$ .

Hence the coefficients of  $\alpha, \beta, \gamma$  separately must be equal on the two sides of the equation: or

$$\sum m \frac{du}{dt} = \sum X$$

$$\sum m \frac{dv}{dt} = \sum Y$$

$$\sum m \frac{dw}{dt} = \sum Z,$$

which equations express exactly the ordinary principle of momentum.

Exactly a similar relation between the three principles of momentum, energy, and relativity is maintained in Einstein's Special Principle of Relativity (1905) which is to be developed later.

## II. THE ADVENT OF ÆTHER THEORY.

It is on the basis of time and space as described above that subsequent physical thought has rested. All motion has been tacitly referred to a Newtonian frame of reference. But the nineteenth century began a new era. Early in that century the idea of a luminiferous æther came into prominence. The question was at once asked, "*What is the velocity of the earth relative to this medium?*"

Arago, sometime before 1818,<sup>1</sup> devised an experiment which he thought would answer that question. On the wave theory of light the index of refraction of a piece of glass is the ratio of the velocity of the incident light to the velocity of the refracted light. If the glass were moving through the æther to meet the incident light, the relative velocity of the light at entering the prism would be greater than if the prism were at rest in the æther. The velocity of the light in the prism should, however, be the same, so Arago thought. Thus the index of refraction and therefore the deviation of the ray of light, should be greater when the prism is moving towards the ray than when it is moving away from the ray or is at rest in the æther. It was such a difference in the deviation that Arago hoped to find when he examined rays of light from stars in different directions, all of which could not make the same angle with the motion of the earth through the æther. The difference corresponding to a velocity of the earth equal to its velocity relative to the sun would have amounted to about one minute of angle, and this was well within the reach of his means of observation, so that he expected to perceive a measurable difference; but he did not. Writing to Fresnel for an opinion on his experiments, he was told that the only place at which his argument might break down was in the assumption

<sup>1</sup> See the letter from Fresnel to Arago on the subject of the experiments "*Annales de Chimie*," 1818.

that the light travelled through the prism at the same relative rate whether the prism was moving or at rest in the æther. Fresnel suggested that the prism did not communicate to the light the whole of its velocity but only a fraction of it. If  $u$  were the velocity of the light in the prism when at rest in the æther, and the prism were set in motion with velocity  $v$  relative to the æther, then Fresnel asserted that the deviation would not be altered at all, if the velocity of the light relative to the prism were now  $(u - v/\mu^2)$ ,  $\mu$  being the ordinary index of refraction; or, in other words, if the prism communicated to the ray not its whole velocity but only the fraction  $(1 - 1/\mu^2)$  of it. This fraction is now commonly spoken of as 'Fresnel's convection-coefficient'.<sup>1</sup>

## 12. MATTER MODIFIED BY ITS MOTION THROUGH THE ÆTHER.

The point of Fresnel's suggestion which is of most significance for us here is that it implies that matter, by its motion through the æther, may be modified in just such a way as to neutralize an effect which would otherwise arise. We shall see later several more instances of hypotheses that have been suggested for the purpose of explaining the non-appearance of effects to which theory seemed to point. It was one of the main objects of the 1905 Principle of Relativity to gather together these separate *ad hoc* hypotheses under a single assumption, and to deduce consequences which do not arise out of the particular and disconnected hypotheses.

## 13. FIZEAU'S VERIFICATION OF FRESNEL'S HYPOTHESIS.

In 1851 Fizeau set out to verify directly Fresnel's suggestion of a convection-coefficient by actually observing the effect of the motion of a stream of water on the velocity of light through it. He devised an experiment in which a beam of light is divided into two parts which traverse two parallel tubes filled with water which can be set in motion with a measurable velocity. A beam of light from the source *a* (Fig. 1)

<sup>1</sup> Larmor, "Æther and Matter," 1900, p. 38, shows that Fresnel's hypothesis is not only sufficient but is necessary if Arago's experimental evidence is universally confirmed.

rendered parallel by the lens *b* is divided into two parts by the half-silvered plate *c*. The two parts are reflected by the mirrors *d* into the two tubes AB, CD which contain the water, and are sent back through the opposite tube by the totally reflecting prism *f*; they are then reunited by the mirrors *d* and the plate *c* and enter the telescope *g*, through which interference fringes are observed.

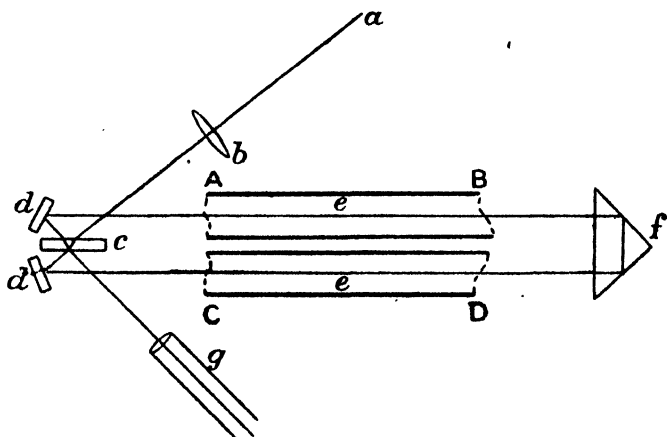


FIG 1.—Fizeau's apparatus as improved by Michelson and Morley.

*Explanation of Diagram.*—*a*, source of light; *b*, condensing lens; *c*, half-silvered plate of glass to divide the beam of light; *d*, *d*, mirrors; *e*, *e*, tubes carrying stream of water in opposite directions; *f*, totally reflecting prism; *g*, observing telescope.

The water in the tubes is then caused to circulate, moving in opposite directions along AB and CD, so that it is moving with one part of the beam of light and against the other. A shifting of the interference fringes is at once observed, which is proportional to the velocity of the water.

Supposing that the velocity of light through the moving water is  $c' + kv$ , where  $c'$  is the velocity when the velocity  $v$  is zero, the times taken by the two parts of the beam to traverse the total path  $l$  in the water are respectively

$$\frac{l}{c' + kv} \text{ and } \frac{l}{c' - kv}$$

the velocity  $v$  being in opposite senses for the two parts.

Thus the retardation of the one beam relative to the other is

$$\frac{l}{c - kv} - \frac{l}{c + kv} = \frac{2lkv}{c^2},$$

neglecting  $(v/c)^2$ .

On measuring the displacement of the interference fringes, Fizeau found that Fresnel's value for the convection-coefficient  $k$ , viz.  $1 - \mu^{-2}$ , fitted his results quite well. The conclusion was confirmed later by Michelson and Morley, who repeated the experiment with modern refinements. The detailed discussion of Michelson and Morley's results will be referred to later (p. 41, §§ 36-8). It will be seen that they are entirely in agreement with the predictions of the principle of relativity.

#### 14. THE SIGNIFICANCE OF FRESNEL'S CONVECTION-COEFFICIENT, AND OF FIZEAU'S VERIFICATION OF IT.

The remarkable experimental verification of Fresnel's brilliant guess considerably strengthened the tendency towards the "stagnant æther" theory. It might have seemed that the hypothesis was a very artificial one, but we must admit that Fresnel's preference for this conception was not without reason. To quote his own words:—

"If we admitted that our globe communicates its velocity to the æther which surrounds it, it would be easy to see why the same prism refracts light in the same manner from whatever direction it comes. But it seems impossible to explain the aberration of the stars on this hypothesis; so far I have not been able to think clearly of this phenomenon except by supposing that the æther passes freely through the earth, and that the velocity given to this subtle fluid is only a very small part of that of the earth; not more than one hundredth part, for example."

The outstanding fact of aberration must, indeed, not be forgotten in considering the various experiments and hypotheses that have been devised to meet the various questions that arise.

Sir George Stokes, a great student of the motion of fluids, tried hard to evolve a theory of the æther which would allow

of it being dragged along with the earth, and at the same time be consistent with the law of aberration. But his theory never received a large amount of support, and was practically abandoned when the establishment of the identity of light and electric disturbances required that the æther should be conceived as much in relation to electro-magnetic phenomena as to optical. But the immediate significance of Fizeau's experiment was that it was thereby definitely shown that the velocity of light relative to a material medium is not a constant depending on the nature of the medium alone, but is also dependent on the velocity of the medium through space, or rather through the æther, which was conceived by Fresnel to be at rest in space.

The effect being a first order one, that is, being proportional to the first power of the velocity, the experiment throws no light on the velocity of the earth itself, since the part of the effect due to this velocity remains practically constant, the observed effect being proportional to the change in the velocity of the water in the tubes relative to the earth. But henceforth it was clear that in respect of its optical properties it must be admitted that a material medium is in some sense modified by its motion.

When at a later date light was identified as an electro-magnetic disturbance this became one deciding factor as between Hertz' theory of electro-magnetic phenomena in moving bodies, and that of Maxwell and Lorentz. Hertz' idea of a moving body was just an extension of the Newtonian idea of a rigid body in motion. The properties of the body were conceived by him to be entirely unchanged when it was set in motion, just as the dynamical mass of a body was unaltered; and among other implications was the complete convection of light, that is, the invariance of the velocity of light relative to the body. In view of Fizeau's experiment it was not possible to maintain this theory, and other experiments soon showed it to be lacking in other respects.

Fresnel's suggestion of the interpenetration of æther and matter therefore assumed greater prominence, and became the basis of the later theory developed by Lorentz and by Larmor.



## 15. THE MICHELSON-MORLEY EXPERIMENT.

The question was, however, by no means regarded as settled by Fizeau's results. In 1881 we find A. A. Michelson returning to it with an attempt at a direct investigation of the difference in the velocity of light relative to the earth in different directions by a method which did not involve propagation through material bodies. It was hoped thus to prevent the possibility of an explanation of the results by means of hypotheses about the modification of matter by its motion through the æther such as Fresnel had advanced in the case of Arago's experiment. The result of this experiment on subsequent thought was so far-reaching that it will be well to describe it in detail.

The experiment was first carried out by Michelson in 1881, and was repeated with greater refinement with the assistance of E. W. Morley<sup>1</sup> in 1887, and again with even greater care by Morley and D. B. Miller in 1905.<sup>2</sup>

Fizeau's experiment, as has been pointed out, being a first order experiment, would only reveal the influence of velocity relative to the earth. In the Michelson-Morley experiment it was proposed to seek for an effect depending on the square of the velocity, and to avoid any question concerning the internal constitution of moving matter. The arrangement was as follows, the figure showing the path of the light relative to the moving apparatus:—

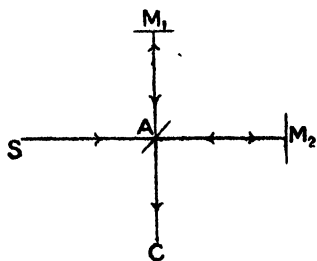


FIG. 2.

A beam of light from a source S was divided by partial reflection at a plate of glass A into two portions travelling along the paths  $AM_1$ ,  $AM_2$ . These two portions were reflected back by mirrors  $M_1$  and  $M_2$ , and, on striking the plate again, a portion of the beam from  $M_1$  is transmitted and brought to interference with a portion of the beam from

<sup>1</sup> "Phil. Mag.," 1887.<sup>2</sup> *Ibid.*, 1905.

$M_2$  reflected along AC. The whole apparatus could be rotated into any position desired.

#### 16. THE THEORY OF THE EXPERIMENT.

Suppose first that  $AM_2$  is in the direction of the earth's velocity relative to the æther, which will be called  $v$ . Then since  $(c - v)$  is the velocity of light relative to the apparatus on the forward journey and  $(c + v)$  is the relative velocity on the return journey, the time taken by the light to travel from A to  $M_2$  and back is

$$\frac{l_2}{c - v} + \frac{l_2}{c + v} = \frac{2l_2c}{c^2 - v^2},$$

where  $l_2$  is the distance  $AM_2$ .

In considering the time taken to go to  $M_1$  and back we have to take a ray whose velocity relative to the apparatus is along  $AM_1$ , which is assumed perpendicular to  $AM_2$ . The relative velocity along this line will thus be  $(c^2 - v^2)^{\frac{1}{2}}$ , and the time taken is therefore

$$\frac{2l_1}{(c^2 - v^2)^{\frac{1}{2}}},$$

where  $l_1$  is the distance  $AM_1$ .

Thus the retardation in time of the former beam relative to the latter when they are brought together again is

$$2 \left\{ \frac{l_2c}{c^2 - v^2} - \frac{l_1}{(c^2 - v^2)^{\frac{1}{2}}} \right\}.$$

If the apparatus is now rotated so that  $AM_1$  comes into the direction of the light,  $l_1$  and  $l_2$  change places, and the retardation is altered to

$$2 \left\{ \frac{l_1c}{c^2 - v^2} - \frac{l_2}{(c^2 - v^2)^{\frac{1}{2}}} \right\}.$$

Thus the change in the retardation is

$$2(l_2 + l_1) \left\{ \frac{c}{c^2 - v^2} - \frac{1}{(c^2 - v^2)^{\frac{1}{2}}} \right\}.$$

#### 17. ALTERNATIVE EXPLANATION.

The following explanation, alternative to that given above, together with fig. 3 which illustrates the absolute path of the rays which interfere, may perhaps make the theory of the experiment a little clearer.

At a certain moment  $t_3$  let  $A_3$  be the position of the moving plate and let us consider the paths of the two parts of the beam which are reunited at that moment. Let  $A_1$  be the position of the plate at the instant  $t_1$  at which that element of disturbance leaves  $A_1$  which, travelling by the path  $A_1M_1A_3$ , arrives at  $A_3$  at time  $t_3$ . Then the time  $t_1$  is found thus:—

$$\begin{aligned} A_1A_3 &= v(t_3 - t_1), \\ c(t_3 - t_1) &= A_1M_1 + M_1A_3 \\ &= 2(l_1^2 + \frac{1}{4}A_1A_3^2)^{\frac{1}{2}} \\ &= \{4l_1^2 + v^2(t_3 - t_1)^2\}^{\frac{1}{2}}, \end{aligned}$$

giving  $(c^2 - v^2)(t_3 - t_1)^2 = 4l_1^2$ ,  
or  $t_3 - t_1 = \frac{2l_1}{(c^2 - v^2)^{\frac{1}{2}}}.$

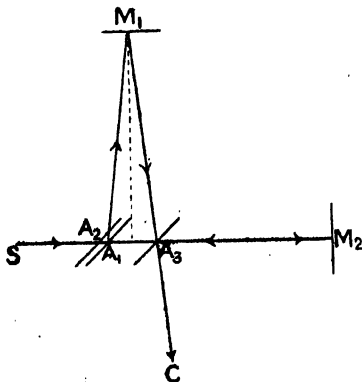


FIG. 3.

In the same way let  $A_2$  be the position of the plate at the moment  $t_2$  of the emission of the element of disturbance which, travelling by the path  $A_2M_2A_3$ , arrives at  $A_3$  at time  $t_3$ .

Then

$$A_3A_2 = v(t_3 - t_2),$$

$$A_2M_2 = c(\tau - t_2),$$

$$M_2A_3 = c(t_3 - \tau),$$

where  $\tau$  is the moment of reflection.

Also  $l_2$  is the distance of  $M_2$  from the position of the plate at time  $\tau$ .

Thus:—

$$l_2 = A_2M_2 - v(\tau - t_2)$$

$$= (c - v)(\tau - t_2),$$

and similarly

$$l_2 = (c + v)(t_3 - \tau),$$

giving

$$\frac{l_2}{c - v} + \frac{l_2}{c + v} = t_3 - t_2,$$

or

$$t_3 - t_2 = \frac{2l_2c}{c^2 - v^2}.$$

Thus the disturbances which are united at  $A_2$  at time  $t_2$  do not leave the plate A simultaneously, but at instants separated by an interval

$$t_2 - t_1 = \frac{2l_1}{(c^2 - v^2)} - \frac{2l_2c}{c^2 - v^2}$$

and therefore will differ in phase by this amount, exactly as obtained above. The question to be answered by the experiment is whether this difference of phase will be altered when the apparatus is turned round.

Looking at the theory from this point of view it becomes necessary to consider whether the directions of the reunited parts of the beam will necessarily be the same. This involves the question of the reflection of a ray of light at a moving mirror. Let us consider this by means of Huygens' principle.

Let AB be a reflector placed at any angle  $\alpha$  with  $SS'$ , and moving with velocity  $v$  in the direction  $SS'$ . Let AX be a plane wave-front incident on the mirror at A at a certain instant; at a small time  $\tau$  later, the unreflected portion of this will have advanced a distance  $c\tau$ , being now part of the line  $X'N$ . In the same time the reflector has advanced a distance  $AA' = v\tau$ ; so that, drawing  $A'B'$  parallel to AB to meet  $X'N$  in C, C is now the point of incidence of the wave-front on the reflector. Thus the reflection is exactly that which would take place at a fixed mirror in the position AC.

In the same way, considering a ray incident on the other side of AB, moving in the opposite direction, we find that  $A'C'$  is the direction of the equivalent fixed mirror.

Thus if the mirror is set at an angle of  $45^\circ$  with  $AS'$  the direction of the reflected part of the beam at the first incidence will be  $AM_1$  where the angle

$$\angle XAM_1 = 2\angle C'AC,$$

and the direction of the transmitted ray after its final reflection will be  $A'Z$  where

$$\angle YA'Z = 2\angle C'A'C.$$

The condition that the reunited rays shall be parallel is that

$$\angle XAM_1 = \angle YA'Z,$$

that is that

$$\angle CAC' = \angle C'A'C,$$

and this is satisfied since the right-angled triangle  $ANC'$  is isosceles.

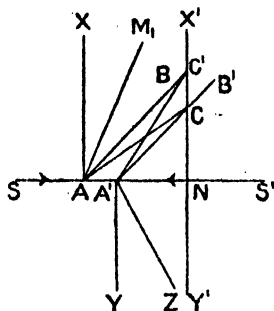


FIG. 4.

## 18. THE RESULT OF THE EXPERIMENT.

Interpreted according to this calculation, if the velocity of

the earth relative to the æther had been as much as a quarter of the velocity of the earth in its orbit, there would have been a displacement of the fringes produced by the interference of the two beams of such magnitude that it could not have escaped observation with the apparatus used. But no trace of such a displacement was found.

Morley and Miller, repeating the experiment in 1905, with still greater refinement, also came to the conclusion that there was no displacement, though they could have observed one due to one-tenth of the velocity of the earth.

### 19. THE FITZGERALD-LORENTZ CONTRACTION.

If, as we seem bound to, we accept the above results as showing the theory which has been given to be inadequate, the hope of determining  $v$  by this means vanishes, and the difficulty remains of reconciling the null result with the hypothesis of an æther which is not convected along with the optical system by the earth.

FitzGerald<sup>1</sup> threw out a suggestion that if the æther can percolate through matter, it may affect the apparatus, and change its dimensions when it is rotated. Such a suggestion had become feasible in view of the adoption of the electromagnetic theory of light, and the growing knowledge of the electrical relations of matter.<sup>2</sup>

If such an effect is to nullify the change in the retardation, we must have, if  $l'_1$ ,  $l'_2$  are the changed lengths of  $AM_1$ ,  $AM_2$ ,

$$\frac{l_1}{(c^2 - v^2)^{\frac{1}{2}}} - \frac{l_2 c}{c^2 - v^2} = \frac{l'_1 c}{c^2 - v^2} - \frac{l'_2}{(c^2 - v^2)^{\frac{1}{2}}}.$$

This is satisfied if

$$\frac{l_2}{l'_2} = \frac{l'_1}{l_1} = \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}},$$

that is, if either arm contracts in the ratio  $1 : (1 - v^2/c^2)^{\frac{1}{2}}$  when turned into the direction of motion, as compared with its length when at right angles to this direction.

<sup>1</sup> See O. Lodge, "Aberration Problems," "Phil. Trans.," 184A (1893), p. 727; also Presidential Address to the British Association, 1913.

<sup>2</sup> E.g. Maxwell's law connecting the index of refraction with the dielectric constant.

FitzGerald's suggestion was not carried further at the time, and it was left to Lorentz to make it indendently,<sup>1</sup> and to show at the same time some plausible reason why a contraction of exactly this amount might be expected.

The theory to which Lorentz was led in his investigations into the optical behaviour of moving bodies is so fundamental to the present subject, and is, if not always in its original form, so generally accepted, that some account of it must now be given.

## 20. THE NULL RESULT INDEPENDENT OF THE MATERIAL CONSTITUTING THE APPARATUS.

Before this is done, however, we may note a fact of great importance for the general significance of the Michelson-Morley experiment. As first carried out, the whole apparatus was mounted on a sandstone block which floated in mercury. In the repetition by Morley and Miller, the distance between the mirrors was intentionally maintained by wooden rods.

Thus if the null effect is to be explained by FitzGerald's suggestion, the contraction of the right amount must automatically take place in two such different materials as sandstone and pine. It is difficult to think of this as a mere coincidence, so that we are naturally led to think that if we accept the contraction hypothesis, whatever explanation of it we may adopt, it must be one that is universal, and inherent in the constitution of matter, even down to the structure of the chemical atom.

## 21. THE INFLUENCE OF THE MICHELSON-MORLEY EXPERIMENT ON THE DEVELOPMENT OF ELECTRICAL THEORY.

The contraction hypothesis of FitzGerald is a second instance of an *ad hoc* addition to existing theory for the express purpose of making it fit the facts. The first instance, cited above, was Fresnel's hypothesis of a convection-coefficient. But it was not possible for the matter to rest at this point. Why should matter contract to exactly the required extent,

<sup>1</sup> "Versuch einer Theorie der Elektrischen und optischen Erscheinungen in bewegten Körpern," Leiden, 1895.

and that, too, in the case of two such different materials as stone and wood? The necessity of answering this question supplied a direct and powerful stimulus to the construction of a theory of matter in which the suggested change in the length of a body is no longer an arbitrary hypothesis, but a necessary consequence.

The result of these efforts was the electron theory of Lorentz and Larmor. In this theory an ideal picture of matter in general is constructed. The special properties characteristic of the particular chemical constitutions of bodies are eliminated from the consideration at the outset; the general laws which govern them are supposed to be common to all kinds of matter. All matter is conceived to be composed of electrons, that is, of small nuclei which are indivisible, and, though this is not absolutely necessary, of identical nature for all kinds of matter. As far as possible the exact nature of these electrons, their size and shape, and the way in which the electric charge is distributed on them, is left out of consideration. These electrons are in motion within the body which they constitute, and set up an electro-magnetic field according to the laws which were first completely formulated by Maxwell.

This field in its turn is supposed to determine the way in which the electrons move. But the general laws which determine the field when the motions of the electrons are given in advance are not sufficient in themselves to do this.<sup>1</sup> To supplement them some further assumption has to be made as to the nature of the reaction of the æther on the electron.

We may see at once how the desire to explain the automatic contraction of a body when set in motion through the æther regulated the choice between the various hypotheses that might present themselves to the mind. Lorentz showed that all that was necessary beyond the equations of Maxwell, to determine the motion of an electron was to assume (i) that its exact size and shape, or, more precisely, the exact distribution of the electric charge, should be given, and (ii) that the whole inertia or kinetic reaction of the electron was due to

<sup>1</sup> For a fuller discussion, see Chapter IV, pp. 51 ff.

the electric field which by its motion it was setting up in the æther. This second assumption is known as the '**hypothesis of purely electro-magnetic inertia**'; it was shown to be a reasonable one by the results of the experiments of Kaufmann in 1901, on the apparent inertia of the free electrons which are supposed to constitute the cathode rays.

Now what was the assumption which Lorentz found it desirable to make about the configuration of the electron? Starting with a desire to prove the reasonableness of the contraction-hypothesis as a universal characteristic of all matter, he was led to assign the same property to the ultimate common element. That is, assuming an electron at rest to have a spherical configuration, an electron moving with velocity  $v$  is supposed to have a spheroidal shape, the whole electron being squeezed together along the direction of the velocity in the ratio  $(1 - v^2/c^2)^{1/2} : 1$ .

The reason for this contraction of the electron is not considered; neither are the agencies which hold the electron together; these are eliminated by means of the purely kinematic assumption. But granting it to exist, Lorentz was able to show<sup>1</sup> considerable reason for anticipating that a body constituted according to his scheme would automatically undergo the required contraction when set in motion.

This, put rather bluntly, is the general result of the work of Lorentz. Of course, there was actually more support for this conception of the electron than the mere fact that it furnished an explanation of the hypothetical contraction. Such, for instance, was the fact that it also gave an explanation of the failure of the experiments of Rayleigh and Brace which sought to find an evidence of double refraction in an isotropic transparent body when set in motion through the æther.<sup>2</sup> But this does not alter the fact that the suggestion of the contracting electron is in reality only throwing the mystery a stage further back. If we accept the suggestion, it

<sup>1</sup> Lorentz, "Proc. Roy. Soc.," Amst., 1904; also, Larmor, "Æther and Matter," 1900. Larmor, on the ground of the theory developed here, demurred to Lord Rayleigh's expectation of any effect (Brit. Ass., Belf., 1902).

<sup>2</sup> See below, p. 56.



will not be long before we go on to ask the further question, 'Why is this so?' This can only be answered by an analysis of the construction of the electron, in terms of its parts, so that we should at once be carried beyond the stage at which the electron is an ultimate element in our thought. At this point, therefore, the assumption must be left until something arises which may throw further light on it, or furnish us with an alternative.

## 22. LORENTZ'S ARGUMENT ANTICIPATES THE PRINCIPLE OF RELATIVITY.

At the present moment it is desired to call attention to the fact that the hypothesis of the contracting electron is nothing more than throwing back on the elementary conceptual constituent of matter the very property which it desired to establish of the objects of experience.

This is in reality an application of what came to be known in 1905 as the Principle of Relativity: that is, it is a result of the assumption that the failure of experiments to discover the motion of the earth relative to the æther is no accident, but is the result of universal relations inherent in the constitution of matter. If Lorentz's or Larmor's theory be true, then the velocity of bodies relative to the æther must for ever remain unknown.<sup>1</sup> This is the point from which the 1905 principle of relativity starts. But the deduction of consequences from this point is independent of any constitutive theory of the nature of matter. Lorentz's theory, as far as it goes, implies the relativity of all its consequences, but many of those results can be obtained as an immediate deduction from the hypothesis of relativity apart from the details of the theory. Not only so, but in regions where the pure electron theory becomes inapplicable, we may say what laws are consistent with the relativity of all phenomena and what laws are not.

The law of gravitation, for instance, in the commonly accepted form, is not consistent with the relativity implied by the electron theory. If matter were completely constituted according to that theory, the law of gravitation would come

<sup>1</sup> See Chapter IV, p. 55.

within the scope of the theory, and we should therefore expect that, even if we could use gravitational phenomena in our investigations, we should still be unable to find any evidence of the earth's motion relative to the æther.

The question is thus suggested :—

(1) Are the observations of astronomy consistent with the hypothesis of relativity in the sense in which we are now thinking of it? Or: Do existing observations give us a means, if only the proper way can be devised, of determining the velocity of the earth through the æther?

(2) If we give up hope of the latter possibility, and grant that in some way the law of gravitation must be included in the scope of the principle of relativity, is this because, contrary to the usual opinion, gravitation is really a product of the electrical constitution of matter, or is the relativity of phenomena in the special sense of which we are thinking, something of wider range and more deep-seated than the electron theory itself?

This fact of gravitation and our incapacity to coordinate it with the other physical qualities of material bodies constitutes perhaps the most important region in which the theory of relativity can go beyond the theories of the constitution of matter. It is just where these theories become insufficient to give a complete account of the mechanism on which the properties of matter depend that the Principle of Relativity becomes of importance as a supplementary hypothesis. The examination of the question as to what light it could throw on the problem of gravitation was no doubt a great stimulus to the revolutionary work of Einstein, which has now gone far beyond the Special Principle of 1905.

## CHAPTER III.

### THE RELATIVITY OF SPACE AND TIME.

23. IN order to arrive at the essential point which Einstein added in 1905 to the argument of Lorentz, we will now begin to examine some positive results of the fundamental assumption that it is impossible to determine the actual velocity of the earth relative to the æther, conceived after the manner of Lorentz as a unique medium which is capable of use as a frame of reference.

It is assumed that the fundamental property of the æther, that of propagating effects with a definite velocity, the same at all points and in all directions, is one of the phenomena with which we are concerned. This implies that we shall assume that the result of the experiment of Michelson and Morley is general; that is, we assume that we are ever to remain in ignorance of any difference between the velocities of light in different directions relative to the earth or, more generally, to any given observer.

Now this means that for all practical purposes we may at any instant suppose the observer to have any velocity we choose. But we may not, of course, attribute to him any arbitrary acceleration; because, as was said at the outset, it is not yet assumed that the relativity extends to the concealment of all motion, but only of a motion which is uniform. For the present, when we speak of an observer we mean an observer who is moving uniformly relative to the æther, so that we may choose a frame of reference in which he is permanently at rest.

### 24. ON THE IDEA OF SIMULTANEITY.

As a result of the fact that an unambiguous conception of time emerges out of the system of dynamics, the idea

of 'simultaneous events at different places' came to be considered as one about which there was no lack of definition.

If we examine the development of the idea, however, we see that it arose gradually in some such way as this. In the first instance, events which were seen by the eye at the same moment were considered to be simultaneous. But in reality the simultaneity was not in the events, but in the observer's consciousness. This is, in fact, the only kind of absolute coincidence about the meaning of which there is no doubt. For as soon as the finite velocity of propagation of light is recognised, it has to be allowed for in estimating the moments of occurrence of events at a distance from the observer. Not only this, but the velocity of the observer must be allowed for if he is thought to be moving.

In doing this the actual observations made are necessarily of what may be called 'coincidences'. Measurement only becomes exact when mental judgment of distances and intervals is eliminated. It is the chief aim of the experimenter to do this as far as possible. If we seek to date some celestial phenomenon, what we observe is a coincidence, that of a light impression falling upon some terrestrial object, as the eye. In allowing for the time of transmission of the light, we enter the realm of theory, and postulate some law such as that of the constancy of the velocity of light. If this is all we postulate, and if we have no other means than that of light signals for setting up a scale of 'simultaneity' for events on the earth, and at distant celestial points, then it will be seen that we are not able to date the event uniquely.

If the conception of a definite and unique æther be adopted, then for an observer at rest in it a criterion of simultaneity can be set up as follows. Let A, B be two points, each of which is supposed to be at permanent rest in the æther, and let a ray of light be emitted from A at an instant  $t_1$ , in the direction AB. Let this ray be reflected back on its arrival at B, the moment of reflection being called  $t_2$ . Let the reflected ray arrive at A again at an instant which we will call  $t_3$ . Then on the ordinary view, A and B being at rest in the æther,  $t$

must be considered as simultaneous with the instant midway between  $t_1$  and  $t_3$ . We may write

$$t_2 = \frac{1}{2}(t_1 + t_3) \quad (1)$$

On the other hand, if A and B have a common velocity  $v$  in the direction AB, and the same process of transmission of a ray of light from A to B and back again is carried out, this will not be so. For if  $l$  represents the constant distance AB and the instants corresponding to  $t_1$ ,  $t_2$ ,  $t_3$  are denoted by  $t'_1$ ,  $t'_2$ ,  $t'_3$ , then we have

$$\begin{aligned} t'_2 - t'_1 &= l/(c - v), \\ t'_3 - t'_2 &= l/(c + v). \end{aligned}$$

These equations lead to

$$\frac{1}{2}(t'_1 + t'_3) + v(t'_3 - t'_1)/2c \quad (2)$$

Thus in this case the instant of reflection is not to be considered as simultaneous with the instant midway between the instants of the start and return of the signal.

Now as long as we have any hope of knowing the velocity of a given point relative to the æther, this fact is in no sense confusing. But if we grant the hypothesis of relativity, and reconcile ourselves to the thought that we shall never know the actual velocity of the points A and B, then we must also face the consequence that we can never know which instant at B may be called, more than any other, 'simultaneous' with a given instant at A. Or we may say that *the phrase 'simultaneous events at different points' has no meaning until the velocity of those points is stated.*

This is one of the main distinctions between the outlook here suggested and that of the earlier thought. *We cannot, granting the fundamental assumption, say that there exists on physical grounds a unique means of ordering phenomena in time regardless of their position.* If on metaphysical grounds we desire to think that there is really a unique and absolute meaning for the time which we measure, yet we must admit that we have no means of knowing whether we are using it or not. It is exactly the same position to which we are brought in respect of the æther. We may assert its existence, but we have no means of identifying it, and the definition of it in

terms of its properties determines it only as one of an infinite number, all of which have identical properties.

## 25. ON THE MEASUREMENT OF LENGTH AND THE CONCEPTION OF RIGIDITY.

If it be granted that we must leave the question of the simultaneity of events an unanswered one, we are also faced with a difficulty about the notion of the distance between two given points of a material body. In the ordinary or Newtonian way of measuring lengths on a moving body, the length of a given line in the body must be defined as the distance between the two points of space which are occupied by the ends of the line *at simultaneous instants*, the conception of distances between the points of space being supposed, like the conception of absolute intervals of time, to be an *a priori* one.

But if we grant that we cannot speak without ambiguity of 'simultaneous instants' until we have chosen an arbitrary body to be the one which is said to be at rest at a particular moment, then it is clear that we cannot, without further definition, speak of the length of a given line in a body until we know what velocity is assigned to it.

It is taken for granted, in the classical treatment of the dynamics of rigid bodies, that the distances between the various parts of a body are invariable in the motion. This conception of an ideal rigid body is fundamental, and space is conceived to be so graduated that a given body occupies a region of exactly the same size and shape at every instant, and that this region is independent of the velocity of the body.

We see now that this conception is breaking down at various points. First we have the suggestion of FitzGerald that every body contracts on being set in motion; then we have the difficulty of defining what we mean by the region occupied by the body at a given instant on account of the difficulty about what we mean by simultaneous positions of different points of the body. Not only this, but we have a further practical difficulty in the fact that if our conception of metrical space is based on that of an ideal rigid body this seems to

react on the measurement of time, since we have the apparent truth that equal distances are traversed by a ray of light in equal intervals of time.

In order to deal with the apparently confused situation so created it is necessary to make an entirely new beginning and to build up afresh our statement of what we mean by the system of measurement of space and time that we employ. In the remainder of this chapter we shall see the position adopted by Einstein in 1905. But his thoroughgoing reconstruction will be described later.

## 26. THE CONSTRUCTION OF ALL POSSIBLE SPACE AND TIME SYSTEMS.

We retain for the present the assumption that we are unable to determine a difference in the velocity of light in different directions.<sup>1</sup> If we did not make some hypothesis of this nature to replace the graduation of space by means of a rigid body and of time by the rotation of the earth, we should have no definiteness in our system of time and space at all.<sup>2</sup> We can order the events which happen in the universe in an infinite number of ways. Suppose, for instance, that we associate with any point of space the quantities  $(x, y, z)$ , and with any instant of time at that point the quantity  $t$ . If we take four quantities  $(x', y', z', t')$ , defined as any arbitrary functions of  $(x, y, z, t)$ , then with any occurrence is associated a set of quantities  $(x, y, z, t)$ , and consequently a set of quantities  $(x', y', z', t')$ . Events may be described, as far as their order is concerned, by relations between these quantities. For example, the motion of a moving point may be described either by saying how  $(x, y, z)$  vary as  $t$  varies, or how  $(x', y', z')$  vary with  $t'$ . Who shall say, or how shall we know, whether the former or the latter is the proper set of quantities associated with our resolution of the history of the physical universe into space and time sequences.

It is only when we begin to speak of the form of the relations which hold between the motions of different points that

<sup>1</sup> Actually this is given up in the new theory.

<sup>2</sup> This is, in fact, the case in the general Principle of Relativity.

the choice of the variables which are used to describe the motions is in any way limited. The Newtonian form of the laws of motion limits us, not to a single set of time and space co-ordinates, but to a group of an infinite number, the relation between any one set  $(x_1, y_1, z_1)$  and any other  $(x_2, y_2, z_2)$  being of the form

$$x_2 = x_1 - ut_1, y_2 = y_1 - vt_1, z_2 = z_1 - wt_1, t_2 = t_1.$$

A fixed point in the second set is a point which in the first set satisfies the relations

$$x_1 - ut_1 = \text{const.}, y_1 - vt_1 = \text{const.}, z_1 - wt_1 = \text{const.};$$

that is, it is a point moving with velocity  $(u, v, w)$ . Here  $(u, v, w)$  are any arbitrary velocities whatever. (See p. 11.)

## 27. THE HYPOTHESIS OF THE CONSTANCY OF THE VELOCITY OF LIGHT.

What now is the corresponding limitation imposed by our attempt to maintain, instead of the Newtonian laws, the fact of the constant and universal velocity of light? This limitation is simply that between any two possible sets of space-time co-ordinates  $(x_1, y_1, z_1, t_1)$ ,  $(x_2, y_2, z_2, t_2)$  there must be such relations that, if a point is moving with velocity  $c$  in the first set, it is also moving with velocity  $c$  in the second set.

This may be set down in mathematical form thus:—

If the moving point changes its space co-ordinates by amounts  $(\xi_1, \eta_1, \zeta_1)$  in an interval of time  $\tau_1$ , and the corresponding quantities in the second system of reference are  $(\xi_2, \eta_2, \zeta_2)$  and  $\tau_2$ ; then as a consequence of the equation

$$\xi_1^2 + \eta_1^2 + \zeta_1^2 = c^2 \tau_1^2$$

we must have

$$\xi_2^2 + \eta_2^2 + \zeta_2^2 = c^2 \tau_2^2.$$

Now, it is a purely mathematical problem to find out for what kind of relation between the two sets of co-ordinates this consequence can and must arise; and it is a problem which is capable of complete solution. The complete solution gives us, of course, certain obvious changes of co-ordinates which we need not discuss; namely (i) the addition of mere constants to  $x, y, z$ , and  $t$ , and (ii) a uniform change in the scale of all four co-ordinates, which amounts to nothing more than a



multiplication of the units of space and time by the same quantity, a process which clearly will not alter the measure of the velocity of a moving point. There is also the obvious possibility of turning the space frame of reference round into any other angular position.

But apart from these there is still an infinite number of ways of changing the co-ordinates. Of these there is an infinite number which are linear transformations,<sup>1</sup> that is, in which each of the new set of co-ordinates can be derived from those of the old set by taking sums of constant multiples of the four co-ordinates of that set. We may show by a simple algebraical calculation that these can all be obtained from a change of variables of the following simple type:—

$$x_2 = \beta(x_1 - vt_1), y_2 = y_1, z_2 = z_1, t_2 = \beta(t_1 - vx_1/c^2). \quad (A)$$

Here  $\beta = (1 - v^2/c^2)^{-\frac{1}{2}}$  where  $v$  is arbitrary.

Suppose now that we consider a point which is at rest in the system of co-ordinates  $(x_2, y_2, z_2, t_2)$ , that is a point for which the space co-ordinates  $(x_2, y_2, z_2)$  have fixed values. Then applying the relation (A) we find that for this point

$$(x_1 - vt_1), y_1 \text{ and } z_1$$

have constant values, that is, the point has a velocity  $v$  parallel to the axis of  $x_1$ , if  $(x_1, y_1, z_1, t_1)$  are considered to be its space-time co-ordinates. This velocity  $v$  is of arbitrary magnitude and direction, since the axis of  $x$  may be taken in any direction we please; so that we may write down a transformation of this type which will cause a point at rest in one system of co-ordinates to have in the new system any velocity that we choose to assign.

28. If we put

$$x_2 = \beta x_1 + \alpha ct_1, y_2 = y_1, z_2 = z_1, ct_2 = \gamma x_1 + \delta ct_1,$$

this gives identically

$$x_2^2 + y_2^2 + z_2^2 - c^2 t_2^2 \equiv x_1^2 + y_1^2 + z_1^2 - c^2 t_1^2,$$

provided that

$$\beta^2 - \gamma^2 = \delta^2 - \alpha^2 = 1$$

and

$$\alpha\beta - \gamma\delta = 0.$$

These conditions lead to

$$\alpha^2 = \beta^2 - 1, \gamma^2 = \delta^2 - 1, \delta^2 = \beta^2$$

so that, if we put  $\alpha/\beta = -v/c$ , we have

$$\beta^2(1 - v^2/c^2) = 1,$$

so that the formulæ for the change of co-ordinates are as stated.

<sup>1</sup> It is possible to generalize all that follows to include the transformations which are not linear.

## 29. THE LORENTZ-TRANSFORMATION.

The transformations of type (A) lie at the heart of the history and early developments of the Principle of Relativity. They are commonly known as Lorentz-Transformations. They were not first arrived at in the way which has been given above. Lorentz was the first to see their significance for the theory of the optical and electrical effects in moving bodies, though he neglected the square of  $v/c$ , and so took  $\beta = 1$ . His results only professed to be accurate as far as the first order in  $v/c$ . The approximation was carried a stage further by Larmor,<sup>1</sup> who showed by the use of the above variables how the second order effects of motion through the æther would be concealed. Einstein was the first to use the transformations in the significance which we have seen them to have.

What we have to do now is to see whether the facts of physical observation are in agreement with the suggestion that any two systems of measurement of space and time which are related to one another by a Lorentz-transformation are equally valid. We have seen that it is consistent with the facts known of the propagation of light through free space. It remains to be seen whether the known facts of electromagnetism and dynamics (including observational astronomy) can be fitted into this framework of space and time without contradiction; these being taken as representing the exact branches of physical science. If this can be done, the result will be that all the facts at our disposal are unable to distinguish between these sets of co-ordinates as valid or convenient measures of space and time.

## 30. ON 'THE DIMENSIONS OF A MOVING BODY.

We are now able to express definitely what is to be assumed of the dimensions of a moving body under the fundamental hypothesis that we have made.

The 'length of any line in a body' is, as was said above, defined to be the distance between two points of space which are occupied simultaneously by the ends of the line. Let us

<sup>1</sup> "Æther and Matter," Cambridge, 1900, pp. 173-7.

suppose that we have a body which is at rest in the system of co-ordinates  $(x, y, z, t)$ . Let the time-space co-ordinates for two particular points of the body be  $(x_1, y_1, z_1, t_1)$  and  $(x_2, y_2, z_2, t_2)$ , and let the corresponding quantities in another system of reference be denoted by the addition of an accent.

Then applying the equation (A) twice and subtracting we have

$$x_2' - x_1' = \beta\{x_2 - x_1 - v(t_2 - t_1)\}, \quad (3)$$

$$y_2' - y_1' = y_2 - y_1,$$

$$z_2' - z_1' = z_2 - z_1,$$

$$t_2' - t_1' = \beta\{t_2 - t_1 - v(x_2 - x_1)/c^2\}. \quad (4)$$

The last equation shows that if we take simultaneous positions of the two points in the second system by putting  $t_2'$  equal to  $t_1'$ , the corresponding times in the first system differ by an amount given by

$$0 = t_2 - t_1 - v(x_2 - x_1)/c^2.$$

But since the body is at rest in the first system,  $x_1$  and  $x_2$  are independent of  $t_1$  and  $t_2$ .

Putting  $(t_2' - t_1')$  equal to the value given above, (4), we obtain from (3) for the difference between the simultaneous values of  $x_2'$  and  $x_1'$  the value

$$x_2' - x_1' = \beta(x_2 - x_1)(1 - v^2/c^2) = (x_2 - x_1)(1 - v^2/c^2)^{\frac{1}{2}},$$

that is the apparent length of a body in the direction of the velocity which it is conceived to have, is less than the apparent length when it is conceived to be at rest, in the ratio  $1 : (1 - v^2/c^2)^{\frac{1}{2}}$ , the dimensions at right angles to the direction of the velocity being the same.

### 31. IS THE FITZGERALD CONTRACTION A REAL ONE?

Naturally there arises in the mind the question whether this apparent contraction is in any sense a real one. To meet this one or two suggestions may be given. Lorentz, working always with the conception of a unique æther, insists that the contraction is a real one, and not, as the position above outlined may seem to imply, a purely subjective change in the system of measurement.

But the question which must be asked is, "What is meant by a real contraction?" In what sense is it possible to speak of

a contraction at all? It must be a contraction relative to something, something which is conceived to have a permanent configuration. This something may be taken to be the conceptual frame of reference, and so long as the frame of reference is a unique one the question as to the existence is one with a meaning. But until the frame is defined this is not so.

The position taken then in regard to the contraction of the apparatus of Michelson's experiment when rotated relative to the earth is this. Let us assume that, when we take a frame of reference in which the apparatus is at rest, the configuration of the apparatus is unchanged when it is turned round so slowly that the effect of centrifugal forces in straining the body may be neglected, as well as the effect due to the altered strain arising from the change in position relative to the gravitational field of the earth. That is, we assume, in accordance with the Principle of Relativity, that an observer on the earth will observe no effect on the apparatus as a consequence of the motion of the earth. This being so, the conclusion drawn from the fundamental hypothesis of the principle by the analysis given above is that, if we refer everything to a frame of reference relative to which the earth is moving, there will actually be a contraction of the apparatus as measured by the co-ordinates of this frame.

But, of course, this real contraction will not be capable of detection; for in order to measure it physically it is necessary to lay alongside of the apparatus another comparison body, and to allow this body to remain alongside and to share the displacement of the apparatus. It will therefore share also in the universal contraction, and the measurements will remain as before.

### 32. CRITICISMS OF THE USE OF LIGHT SIGNALS.

The objection is often raised against the argument of the preceding sections that the introduction of this fundamental assumption of the constant velocity of light for the purpose of controlling the time standards at different places is in reality a very artificial proceeding. This objection should be considered carefully before going any further.

The first question that may be asked is, "Have we any right to suppose that there is no means of communicating between distant points other than, and independent of, the transmission of light. If such a means existed it might be used to set up a standard of simultaneity which is different from that given by the equation (2), p. 30."

It used often to be suggested, for instance, that gravitation is propagated instantaneously, or, at any rate, with a velocity enormously greater than that of light. In view, however, of the new theory of gravitation to be described later it will not be worth while discussing this here. The whole of the present argument must be taken as preparing the way for the broader outlook.

A second form of the same objection is that in practice our measure of time is controlled by the dynamical motion of the earth. This amounts, as a matter of fact, to saying that the time which we use in practice is the uniquely defined time of the Newtonian dynamics. But, as we shall see later, the classical dynamics cannot be considered any longer as absolute in its accuracy. This is obvious merely from the fact that we know that moving charged bodies do not obey the Newtonian laws of motion; the deviation has been actually observed in the case of the cathode rays and the  $\beta$ -rays. We shall hardly expect therefore, that with the modern view that matter is largely, if not wholly, of electrical constitution, the dynamics of material bodies will conform *exactly* to the simple scheme propounded by Newton. We are almost prepared, in fact, to believe that the exact dynamics of matter will be a very complicated matter. Abraham, for example, dealing with the mechanics of the electron, obtains a formula<sup>1</sup> for the apparent mass of a single electron which makes it appear almost hopeless to think of further progress in the direction of evaluating the mass of a much more complicated system.

Now, as we shall see later<sup>2</sup> just at this point the Principle of Relativity was able to step in and indicate a general modification of the usual formulæ of dynamics which renders

<sup>1</sup> "Theorie der Elektrizität," vol. ii, 2nd edition, Leipzig, 1908, p. 180.

<sup>2</sup> Chapter V.

them consistent with the standards of time and space set up in the preceding sections. The modifications required are everywhere of the order of the square or some higher power of the ratio of the velocity of the body to that of light, and if these quantities are neglected the formulæ reduce to the ordinary Newtonian form.

These modifications were found to be in strict accordance with the direct evidence available from experiments on the motion of the free negative electrons in the cathode and  $\beta$ -rays,<sup>1</sup> which at present are the only moving systems in which the velocities are sufficiently great to render the deviation from Newtonian mechanics appreciable.

Further, there is strong indirect support in the null results of the experiments of Rayleigh and Brace, Trouton and Noble, which cannot be explained by a combination of existing electrical theory with Newtonian dynamics.<sup>2</sup>

The objections raised are therefore such as can only be maintained by showing either that there are some experimental facts which are not reconcilable with the hypothesis of constant light velocity, or else by showing that an unnecessary complication is introduced into our thought by making this assumption.

### 33. THE VELOCITY ADDITION FORMULA.

If we consider a moving point which is displaced in time  $\delta t$  through distances  $\delta x$ ,  $\delta y$ ,  $\delta z$  parallel to the three co-ordinate axes, the corresponding displacements when measured in the system  $(x', y', z', t')$  can be obtained by applying the equations (A) twice for the positions at the beginning and end of the interval  $\delta t$ . On subtraction we get

$$\delta x' = \beta(\delta x - v\delta t), \delta y' = \delta y, \delta z' = \delta z, \delta t' = \beta(\delta t - v\delta x/c^2) \quad (5)$$

From these, dividing  $\delta x'$ ,  $\delta y'$ ,  $\delta z'$  by  $\delta t'$ , we get expressions for the velocities as estimated in the new frame of reference:—

$$u'_x = \frac{u_x - v}{1 - vu_x/c^2}, u'_y = \frac{\beta u_y}{1 - vu_x/c^2}, u'_z = \frac{\beta u_z}{1 - vu_x/c^2} \quad (a)$$

<sup>1</sup> Vide § 52, p. 65.

<sup>2</sup> §§ 47-9 *infra*.

If we put  $u$  equal to zero in these equations,  $u'$  becomes equal to  $(-v)$  in the direction of the axis of  $x$ .

Conversely since, as we have seen, the formulæ (A) can be reversed by simply changing the sign of  $v$ , if  $u'$  is zero,  $u$  is equal to  $v$  in the positive sense.

Thus a point at rest in the first system of co-ordinates and a point at rest in the second system have a velocity  $v$  relative to one another.

#### 34. RELATIVE VELOCITY.

In the ordinary way of speaking the relative velocity of two points is defined as the difference of their velocities, this being the same for all frames of reference. But as a consequence of the equations (a) this definition would not give us a determinate value for the relative velocity. We therefore substitute the following definition: *the velocity of a point P relative to a point Q is the velocity of P in a system of co-ordinates in which Q is at rest.*

Such a system can always be chosen; for if we take an arbitrary system, and the velocity of Q in this system is  $v$  then taking the axis of  $x$  in the direction of this velocity, and applying the transformation (a), the velocity of Q in the new system is, as we have seen, zero.

Thus, for example, if we have three points A, B, C, of which B has a velocity  $u$  relative to A, and C has a velocity  $v$  relative to B in the same direction as  $u$ , then the velocity of C relative to A is not  $(u + v)$ , but  $(u + v)/(1 + uv/c^2)$ .

#### 35. THE VELOCITY OF LIGHT AS A CRITICAL VELOCITY.

From the relations above, (5), p. 39, we have identically

$$\delta x'^2 + \delta y'^2 + \delta z'^2 - c^2 \delta t'^2 = \delta x^2 + \delta y^2 + \delta z^2 - c^2 \delta t^2,$$

or

$$\delta t'^2(u'^2 - c^2) = \delta t^2(u^2 - c^2); \quad (6)$$

hence if  $u$  is less than  $c$ , so is  $u'$ ; that is, if the velocity of a moving point is less than the velocity of light in any system of reference, it is also less in any other, provided always that the velocity  $v$  is also less than  $c$ .

An example of this may be given in reference to an objection that has often been raised to the point of view of the Principle

of Relativity. It is suggested<sup>1</sup> that whereas the analysis of the principle indicates that relative velocities greater than that of light are physically impossible, yet there is an actual possibility of observing in nature a relative velocity considerably greater. We may imagine, for instance, two  $\beta$ -particles shot out by a radio-active body in opposite directions with velocities, each equal to nine-tenths of that of light; and here is an actual difference of velocity equal to  $1.8c$ . This is quite true, and the Principle of Relativity has nothing to say against it. The principle maintains that a velocity greater than  $c$  *relative to the observer* cannot be observed; this is in accordance with the facts in the case in question. Further, the relative velocity of two bodies, as defined above, cannot be greater than  $c$ ; this also is in agreement with the example chosen; for if in the expression  $(u + v)/(1 + uv/c^2)$  we put both  $u$  and  $v$  equal to  $.9c$ , we get, not  $1.8c$ , but  $1.8/1.81c$ , which is certainly less than  $c$ .

The position is not that velocities greater than  $c$  are not conceivable, but that real bodies become illusory in observation if they are conceived to be moving faster than light. We shall also see later that the electrical constitution of matter seems to indicate that a body would suffer dissolution if it were accelerated so that its velocity were made greater than  $c$ . Not only so, but the energy of a body tends to an infinite value as its velocity is increased towards  $c$ , so that consistently with our ordinary conceptions of energy, it would not be possible to set a body in motion from rest with this velocity. Of course, there is nothing to hinder us if we please from thinking of bodies which have a real configuration for a velocity which is greater than  $c$ ; but such bodies would suffer dissolution if they were caused to move with a velocity less than  $c$ , and an infinite amount of energy would be used in the retardation.

### 36. THE FIZEAU EXPERIMENT AND FRESNEL'S CONVECTION-COEFFICIENT.

We may now consider the way in which the above considerations give an account of the result of Fizeau's experiment.

<sup>1</sup> Vide Soddy, "Nature," November, 1913.



We suppose that a light disturbance travels through a refracting substance with a relative velocity  $c/\mu$ ; that is, this is the velocity in a frame of reference in which the refracting material is at rest. Suppose now that we change to a frame of reference in which the material has a velocity  $v$  in the direction of propagation of the light.

Then, applying the addition equation, the velocity of the light disturbance in this frame of reference is

$$\frac{\frac{c}{\mu} + v}{1 + \frac{c}{\mu} \frac{v}{c^2}} = \frac{c}{\mu} + v \left( 1 - \frac{1}{\mu^2} \right) + v^2 \dots$$

Thus, neglecting  $v^2$ , we arrive at once at the convection-coefficient  $(1 - \mu^{-2})$ .

In the repetition of Fizeau's experiment by Michelson and Morley the figure arrived at for this coefficient was  $\cdot 442 \pm \cdot 02$ .<sup>1</sup> Taking the known value of  $\mu$  for sodium light the theoretical value is  $\cdot 438$ . Thus the agreement is satisfactory enough.

### 37. THE DOPPLER EFFECT.

In the same way, we may give also a simple account of the Doppler effect.

If light, of period  $\tau$  to a given observer, is travelling along the axis of  $x$ , any component of the disturbance may be taken as proportional to a harmonic function

$$\sin \left\{ \frac{2\pi}{\tau} (t - x/c) + a \right\}$$

where  $x$  and  $t$  are space-time co-ordinates in which the observer is at rest.

If now we change to a system in which the observer has velocity  $(-v)$  by putting

$$x = \beta (x' + vt'), \quad t = \beta (t' + vx'/c^2)$$

we have

<sup>1</sup> "Amer. Journ. of Sci.," 31, 1885. The figure given in this paper is  $\cdot 434$ , but there appears to be an arithmetical error, the data given leading to  $\cdot 442$ .

$$\begin{aligned}\frac{2\pi}{\tau}(t - x/c) &= \frac{2\pi\beta(1 - v/c)}{\tau}(t - x/c) \\ &= \frac{2\pi}{\tau'}(t' - x'/c)\end{aligned}$$

where

$$\begin{aligned}\frac{\tau}{\tau'} &= \beta(1 - v/c) \\ &= 1 - v/c \text{ neglecting } (v/c)^2.\end{aligned}$$

Here  $\tau'$  is the period as seen by an observer at rest in the new frame of reference. The result indicates that the effect of a motion with velocity  $v$  opposite to that of the light shortens the apparent period in the ratio  $(1 - v/c) : 1$ .

### 38. THE CONVECTION-COEFFICIENT IN A DISPERSIVE MEDIUM.

In treating of the convection-coefficient in § 36,  $\mu$  was defined by the fact that  $c/\mu$  was the velocity of the disturbance when the medium was treated as stationary. In a dispersive medium  $\mu$  is a function of the period of the light. The value to be taken in the formula obtained is, clearly that corresponding to the period  $\tau$  as seen by the observer to whom the medium is stationary. But the period of the light as observed in Fizeau's experiment is  $\tau'$ , that which is apparent to the observer relative to whom the medium is moving with velocity  $v$ .

By the same method as in the last section we obtain

$$\tau' = \tau \left( 1 - \frac{v}{c/\mu} \right).$$

If  $\mu'$  is the index corresponding to  $\tau'$

$$\begin{aligned}\mu' - \mu &= (\tau' - \tau) \frac{\partial \mu}{\partial \tau} \\ &= - \frac{v\mu\tau}{c} \frac{\partial \mu}{\partial \tau}.\end{aligned}$$

Thus

$$\frac{c}{\mu} = \frac{c}{\mu'} \left( 1 - \frac{v\tau}{c} \frac{\partial \mu}{\partial \tau} \right).$$

$$\text{Hence } \frac{c}{\mu} + v \left( 1 - \frac{1}{\mu^2} \right) = \frac{c}{\mu'} + v \left( 1 - \frac{1}{\mu'^2} \right) - \frac{v\tau}{c\mu'} \frac{\partial \mu}{\partial \tau}.$$

In the last term we may neglect the difference between  $\mu$  and  $\mu'$ .

The additional term in the coefficient of  $v$  increases the theoretical value of the convection-coefficient in the case examined by Michelson and Morley to '451, but this is still well within the limits indicated by their experiments, namely, '442  $\pm$  '02. Zeeman<sup>1</sup> has more recently performed some further very careful experiments with light of different colours with a view to confirming the formula last given. He has succeeded in showing a dispersive effect agreeing very closely with the above formula.

In the following table, the values of the convection-coefficient are given; in the first column, according to Fresnel's formula ( $\epsilon_F$ ), in the second, according to the formula here given ( $\epsilon_L$ ), and, in the last, the value experimentally found.

Wave-length in Å°	$\epsilon_F$	$\epsilon_L$	$\epsilon_{EXP.}$
4500	0'443	0'464	0'465
4580	0'442	0'463	0'463
5461	0'439	0'454	0'451
6870	0'435	0'447	0'447

### 39. THE CONTRACTION IN VOLUME OF A MOVING BODY.

It follows at once from the formula obtained above that if a body, all of whose points have the same velocity, has a volume  $V_0$  when it is considered to be at rest, and volumes  $V$ ,  $V'$  when its velocities are  $u$  and  $u'$  respectively, then

$$V = V_0(1 - u^2/c^2)^{\frac{1}{2}}, \quad V' = V_0(1 - u'^2/c^2)^{\frac{1}{2}}.$$

Now if  $u'$  is the velocity obtained from the velocity  $u$  by the equations (A), we find on substitution from the addition formula that

$$1 - u'^2/c^2 = (1 - vu_x/c^2)^{-2}$$

<sup>1</sup> "Amst. Proc." XVII. (1914), p. 445; XVIII. (1916), p. 398 and p. 1240.

which leads to  $\frac{(1 - u^2/c^2)^{\frac{1}{2}}}{(1 - u'^2/c'^2)^{\frac{1}{2}}} = \beta(1 - vu_x/c^2),$

so that  $V/V' = \beta(1 - vu_x/c^2). \quad (7)$

This equation gives us the contraction in a volume which is not at rest in either frame of reference. It will be of importance when we come to consider later whether the measure of an electric charge is an absolute or relative magnitude (p. 49).

## CHAPTER IV.

### THE RELATIVITY OF THE ELECTRO-MAGNETIC VECTORS.

#### 40. THE DEFINITIONS OF ELECTRIC AND MAGNETIC INTENSITIES.

IN the ordinary treatment of electrical theory the definitions given of the electric and magnetic intensities at a point are somewhat as follows:—

The 'electric intensity' is the force which would be exerted per unit charge on a small charged body placed at rest at the point.

The 'magnetic intensity' is definable from the law that the force per unit charge on a small moving charged body is equal to the sum of the electric intensity and the vector product<sup>1</sup> of the velocity and the magnetic intensity.

An alternative definition of the magnetic intensity is often given as the force per unit pole on a magnetic pole placed at the point. As this introduces the non-existent isolated magnetic pole, it will be as well to confine the discussion to the former definition.

We see that, apart from the fact that the measurement introduces a reference to the mechanical category of 'force,' the definitions involve the knowledge of the velocity of the small test body. But we have already learned to look upon the velocity of a body as a quantity which, physically, is so far indeterminate. To assign a definite value to it we have to decide which of all possible frames of reference we will adopt. Thus the definitions of the electric and magnetic intensities are relative ones; their values are 'relative' in the same sense as the measures of space and time.

We need therefore to add to the relation (A) (p. 34), a relation between the values assigned to the electric and magnetic

<sup>1</sup> See note on vectors, p. 1.

vectors in two different systems of reference. We cannot obtain this from the definitions given just now, because they involve the conception of 'force,' and we have not yet seen whether this is to be looked upon also as a relative or absolute quantity. We shall see as a matter of fact when we come to the consideration of dynamical quantities that we are bound to treat it as relative.

In order therefore to obtain the required relations we must adopt the same method as that used above in respect of space and time, and consider the measures of these quantities in the light of the laws in which they occur, the laws which express the uniformities which we believe on experimental grounds to hold between electrical phenomena.

These laws are the equations of the electro-magnetic field. We will take them in the form adopted by Lorentz,<sup>1</sup> namely,

$$\frac{1}{c} \left( \frac{\partial \mathbf{e}}{\partial t} + \rho \mathbf{u} \right) = \text{curl } \mathbf{h} \quad . \quad . \quad \text{I}$$

$$0 = \text{div } \mathbf{h} \quad . \quad . \quad \text{II}$$

$$- \frac{1}{c} \frac{\partial \mathbf{h}}{\partial t} = \text{curl } \mathbf{e} \quad . \quad . \quad \text{III}$$

$$\rho = \text{dive } \mathbf{e} \quad . \quad . \quad \text{IV}$$

where  $\mathbf{e}$  and  $\mathbf{h}$  are the electric and magnetic intensities, and  $\rho$  is the density of charge.

The fifth equation of Lorentz's theory,

$$\mathbf{F} = \mathbf{e} + [\mathbf{uh}]/c$$

where  $\mathbf{F}$  is the 'mechanical force' per unit of charge moving with velocity  $\mathbf{u}$ , embodies the definitions given above, and supplies us with a means of connecting up with mechanical theory, but gives us no criterion about the purely electrical magnitudes, the mechanical categories now occupying a derivative place.

#### 41. THE INVARIANT FORM OF THE FUNDAMENTAL EQUATIONS.

The first thing that may be noticed about the form of the equations I-IV, is that if we change the scale on which we

<sup>1</sup> The units here used are those of Lorentz. See "Theory of Electrons," Leipzig, 1909, p. 5.

measure  $\mathbf{e}$  and  $\mathbf{h}$ , we alter the scale on which we measure  $\rho$  in the same ratio. Thus it becomes possible to limit the consideration to such changes of the variables as leave the charge carried by any particular element of volume unaltered.

The fundamental hypothesis of electrical theory which takes the place of the Newtonian conception of 'conservation of mass' is that of '*conservation of electric charge*'. We proceed to examine, therefore, whether it is possible to make such changes in the variables  $\mathbf{e}$  and  $\mathbf{h}$ , when we change the space-time co-ordinates according to the transformation (A), that *the form of the equations I-IV may be unaltered* and at the same time *the charge carried by any particular element of volume may be unaltered*. The Principle of Relativity sprang from the discovery that it is possible to satisfy these requirements.<sup>1</sup>

The following are the equations which give us the necessary changes in the vectors  $\mathbf{e}$  and  $\mathbf{h}$  :—

$$\left. \begin{aligned} e'_x &= e_x, e'_y = \beta(e_y - v h_x/c), e'_z = \beta(e_z + v h_y/c) \\ h'_x &= h_x, h'_y = \beta(h_y + v e_x/c), h'_z = \beta(h_z - v e_y/c) \end{aligned} \right\} \text{(B)}$$

If we take the equations of Lorentz and make the changes of variables given by (A) and (B), we arrive directly at the equations

$$\frac{1}{c} \left( \frac{\partial \mathbf{e}'}{\partial t'} + \rho' \mathbf{u}' \right) = \text{curl}' \mathbf{h}' \quad \text{I'}$$

$$0 = \text{div}' \mathbf{h}' \quad \text{II'}$$

$$-\frac{1}{c} \frac{\partial \mathbf{h}'}{\partial t'} = \text{curl}' \mathbf{e}' \quad \text{III'}$$

$$\rho' = \text{div}' \mathbf{e}' \quad \text{IV'}$$

where  $\mathbf{u}'$  is a velocity whose components are

$$\left. \begin{aligned} u'_x &= \frac{u_x - v}{1 - v u_x/c^2} \\ u'_y &= \frac{\beta u_y}{1 - v u_x/c^2} \\ u'_z &= \frac{\beta u_z}{1 - v u_x/c^2} \end{aligned} \right\} \text{(a)}$$

and  $\rho'$  is defined by the equation

$$\rho' = \beta \rho (1 - v u_x/c^2) \quad \text{(C)}$$

<sup>1</sup> A. Einstein, "Ann. der Phys.," 17, 1905, pp. 891-7.

In order to see the significance of these quantities  $\mathbf{u}'$  and  $\rho'$ , we must recall one or two of the immediate consequences of the change in the manner in which space is measured from the co-ordinates  $(x, y, z, t)$  to  $(x', y', z', t')$ .

We have seen, (p. 39) that the equations (a) are exactly those which connect the estimates of the velocities of a moving point as measured in the two systems respectively.

We have also seen, (7), p. 45, that if  $\delta V$  is the volume of a small region of a body moving with velocity  $\mathbf{u}$  as measured in  $(x, y, z, t)$ , and  $\delta V'$ ,  $\mathbf{u}'$  are the corresponding volume and velocity as measured in  $(x', y', z', t')$  then

$$\delta V' = \delta V / \beta (1 - v u_x / c^2).$$

If we put this equation and (C) together we obtain

$$\rho' \delta V' = \rho \delta V.$$

Thus the equation (C) is exactly that which is required to express that the *charges in the corresponding elements of volume are the same*.

The equations (A), (B), (C) then effect an exact transformation of the fundamental equations I-IV into equations of the same form, and are consistent with a unique value being assigned to any element of charge. We say that '*the charge is an invariant*'.

#### 42. THE CONSEQUENCE OF THIS TRANSFORMATION.

The result of the exactness of this transformation is that, so long as we are dealing with phenomena of which the sole laws are the field equations in the form which we have taken, we shall not be able to discriminate between one set of space-time co-ordinates and another, as the only one relative to which the laws have the required form. In other words, remembering that as we know it, the æther is simply the frame of reference relative to which these equations hold, as long as we are dealing with phenomena which are entirely governed by these laws, we can never expect to identify the æther. Unless we can light on some measurable phenomena which are not subject to these field equations alone, the æther remains an unknown frame of reference, exactly as the frame of reference for the laws of dynamics is indeterminate.



Lorentz remarks that it is really a matter of personal predisposition whether this leads a physicist to deny the existence of the æther or to the more agnostic position which merely says that "the æther is real but will always be hidden from our knowledge". This may be so, but few to whom the means of the propagation of electrical effects, and, indeed, the whole structure of matter is a question of the first interest, will feel content to take either of these positions. If the æther, in the ordinary sense of a stationary medium, does not exist, what is it that conveys the waves of light through space? If it does exist, is it a pure accident that its laws are of such a form that it must for ever remain concealed; or is it that we have naturally lighted first on those aspects of its working which, as being independent of motion through it, would be most readily found by an observer who did not know his velocity through it?<sup>1</sup>

In any case, the impression made by the situation is that there is a break in our knowledge. The Principle of Relativity represents an attempt to see what are the conclusions to be drawn from an admission of our inability to identify the æther, in the hope that it may throw light on the question where we are to look for the way out of this difficulty. In order to see at what point the Principle is an independent hypothesis, something which is not contained in a purely mathematical result arising out of the particular form of a set of equations, we must go a little further, and consider the way in which the electron theory professes to give an account of the structure of matter.

#### 43. THE METHOD OF COINCIDENCES IN EXPERIMENTAL OBSERVATION.

The fact that we cannot have a unique measure of the electric and magnetic intensities may perhaps be a little more readily accepted if we consider that in practice we never really profess to measure them directly. When we, for example, measure an electric potential by means of an electrometer, the only actual observation is one of a configuration of relative equi-

<sup>1</sup> This is suggestive of Prof. Eddington's observation in regard to the general principle of relativity—"Space, Time, and Gravitation, p. 201: "The mind has but regained from nature that which the mind has put into nature."

brium. We observe the *coincidence* of a pointer or of a spot of light with a certain mark on a scale. Only by eliminating the uncertain factor of human judgment of an interval of space and time is it possible to make accurate measurements. We make our instruments do all the observing for us. Then if we want to interpret the indications of the instrument in terms of the values of the strength of an electric field, or of an electric current or the like, we have to do so by means of exactly those theoretical equations which we have seen to have a purely relative significance.

Now it is a fundamental property of the systems of co-ordinates that we have been considering, that events which are simultaneous and coincident in space at a given instant in a given set of co-ordinates are also simultaneous and coincident in any other of the sets of co-ordinates. To a given set of values of  $(x, y, z, t)$  there is only one value of each of the co-ordinates  $(x', y', z', t')$ , and these values do not depend upon any particular circumstances associated with the place or instant. Thus if we read an electrometer, the coincidence of pointer and a certain mark on the scale does not depend on the frame of reference that we are using. But the significance that we attach to the observation does do so. The quantitative description of the state of an electrical system depends on the frame of reference used; the nature of this dependence is limited by the fact that the laws of the field are to be independent of the choice of the frame of reference, and this means that the estimates of electric and magnetic force for two different frames of reference will be related in exactly the way represented by the equations (B) which are necessary for the preservation of the form of the laws.

#### 44. THE INCOMPLETENESS OF PURELY ELECTRICAL SCHEMES OF THE CONSTITUTION OF MATTER.

The foundation of the electron theory of matter is the scheme of equations which has been considered in the preceding sections. But there are certain other hypotheses that have to be added to those general laws.

The first is that of '*the atomic nature of electricity*'. The negative '*electrons*' are supposed to be small discrete nuclei,

of identical structure and charge. It is not necessary here to go into the evidence for this. The magnitude of these nuclei is not definitely known; it is only possible for us to construct tentative theories as to their exact structure. For certain purposes we may be able to treat them as points, for others we may need to adopt some provisional conception of their extent and constitution. For the present so little is known about the properties of positive electricity that we must be content to admit that it is by no means certain that a similar statement of the identity of the elementary quanta of positive electricity is possible.

The picture which we have of the constitution of a portion of matter is one in which these electrons are grouped and held together in a permanent configuration; not in a static state, but in a state of motion; the motion of each electron is governed by certain laws of which the theoretical equations that have been already referred to are a *partial* expression.

As they stand it is impossible for them to be the complete scheme of laws. For it is easily seen that they are not mathematically sufficient to determine, from a knowledge of the whole field at any given instant, the way in which the field and the distribution of charge will vary through all time. The equations II and IV (p. 45) are not independent of I and III,<sup>1</sup> so that we really have only two vector equations connecting the three vectors  $\mathbf{e}$ ,  $\mathbf{h}$ , and  $\mathbf{u}$ ,  $\rho$  being connected

<sup>1</sup> From III, since  $\text{div curl } \mathbf{e} = 0$  (see note on Vector Analysis, p. 1), we have  $\frac{\partial}{\partial t} \text{div } \mathbf{h} = 0$ , showing that at a given point  $\text{div } \mathbf{h}$  has a constant value. If at any preceding time there was no field, this constant value must be zero. Thus Eqn. II is deduced from III. From I and IV we obtain

$$\frac{\partial}{\partial t} (\text{div } \mathbf{e}) + \text{div } (\rho \mathbf{u}) = 0,$$

or

$$\frac{\partial \rho}{\partial t} + \text{div } (\rho \mathbf{u}) = 0.$$

This is what is commonly known as the equation of continuity or of conservation of electricity. Integrated through a closed volume it gives—

$$-\frac{\partial}{\partial t} \iiint \rho dV = \iint \rho \mathbf{u} \cdot d\mathbf{S},$$

that is, the rate of loss of electricity included in volume is exactly accounted for by the outward flow over the surface; there is no creation or destruction of electricity.

with  $\mathbf{u}$  by the condition of 'the conservation of electricity,' which requires that the change in the distribution of electricity takes place entirely by convection.

We cannot therefore pretend that our scheme of the structure of matter is even theoretically complete until we have added to the general equations some further condition. It is just at this point that the structure of the electron has to be taken into account. If we choose to assume that we may neglect its dimensions, and treat it as a point charge, this is a sufficient condition.<sup>1</sup> The quantity  $\rho$  is then everywhere zero, and the solutions of the equations which represent the field are such as are restricted to possess singularities of a certain definite type. But as a matter of fact this is not a possible way of building up a practical theory because it introduces mathematical difficulties in the very simplification of ignoring the dimensions of the electron. The specification of the exact way in which the electric and magnetic forces become infinite in the neighbourhood of a point-electron is a matter of considerable difficulty, and the only way of avoiding this is to assume the electron to have a structure of finite size, and to assume some definite relation between the distribution of charge through its volume, and the quantities  $\mathbf{e}$ ,  $\mathbf{h}$ , and  $\mathbf{u}$ . Such a hypothesis will be purely empirical; it must be chosen according to the indications which we have of facts that are not explained by the general equations.

So the question arises: "What light is thrown by experiment on the structure of the electron?" This question has been rather obscured recently by the more tangible one of the structure of the atom as a group of electrons; but this has only become possible because, on any theory of the constitution of the electron, its extreme minuteness does allow of it being treated for some purposes as a material particle of definite mass and charge, so that it is amenable to ordinary mechanical treatment.

It would seem that if the structure of the electron can be ignored in problems of the atom, it would *a fortiori* be so in the discussion of the grouping of atoms to form matter. But

<sup>1</sup> As in Larmor's "*Æther and Matter*".

in the particular problem which we are considering, namely, the failure of experiment to determine uniquely the frame of reference which we call the æther, this is not so. For the preceding analysis of the fundamental equations has shown us that this failure arises partly because the equations of the electric field have a form which is invariant under a certain transformation. We are thus able to form any number of pictures of a given field, each using a different frame of reference, and each subject to the same fundamental equations. But if we wish to extend our picture so as to take in not only the æthereal field but the whole of the matter with which it is in interaction, it is necessary that any further hypotheses, introduced to supplement the equations of the field to obtain a complete scheme for the determination of the configuration and motion of the various parts of the matter, must also be of such form that they too are independent of the particular frame of reference that is adopted.

In order to illustrate this point we may consider a little more in detail the explanations offered by Lorentz of the failure of specific experiments, as representing the point of view of one who thinks in terms of the actual existence of a unique stationary æther. (Cf. Chapter II.)

#### 45. THE EXPERIMENT OF MICHELSON AND MORLEY.

We have already seen that the failure of this experiment to discover any trace of a motion relative to the æther can be completely accounted for by the hypothesis of a contraction of the apparatus. The magnitude of this contraction is exactly that which is apparent when, the body being supposed unaltered, the frame of reference is changed according to the transformation of Einstein. What reasons are given for supposing that the body will actually undergo this contraction when its velocity relative to a fixed frame of reference is changed?

The argument, relieved of the mathematical details, is as follows: When a body is at rest in the æther each electron within it has a specific motion determined according to certain laws. If we know these motions completely we can

by means of the space-time correlation build up a picture of another system which is moving as a whole with velocity  $v$  relative to the same æther; in this the motion of any particular electron is obtained from that of the corresponding electron in the original body by merely turning the relations between  $(x, y, z)$  and  $t$  that express the way in which it moves into relations connecting  $(x', y', z')$  with  $t'$ . If this is done it is clear that the configuration of the body as a whole in the new picture is exactly that of the original body contracted in the ratio  $1 : (1 - v^2/c^2)^{1/2}$ , as on page 36. But, the argument goes on, the electro-magnetic field of the electrons in the new body satisfies the same fundamental laws as that of the original body, and is therefore equally capable of existing. Thus if these were the only laws in question we should be fairly justified in concluding that the original body would automatically change into the one of which we have so formed a picture when it is set in motion with the velocity  $v$ .<sup>1</sup>

#### 46. INCOMPLETENESS OF THE ARGUMENT FOR THE FITZGERALD CONTRACTION.

It is clear that there are two gaps in this argument. First, there is the obvious restriction that it must be assumed that, provided that the acceleration of the body is sufficiently slow, so as not to produce mechanical distortion, change of temperature or other disturbances, there is a *unique* configuration for a given grouping of electrons moving as a whole with a given velocity relative to the æther. This does not seem to offer any serious difficulty, though it must be remembered that the external configuration of a material body must be thought of as a statistical one; that is, in view of a probable kinetic constitution, to a fixed external configuration will correspond a multitude of rapidly changing distributions of electrons in the interior; but this will not be considered here. The more important point for discussion is that the correlation which has been shown to exist in the case of the 'field equations,' must be *assumed* to hold for all the other relations which play a part in determining the motions of the electrons.

<sup>1</sup> Cf. also "Æther and Matter," p. 176.

It has been remarked that it is enough for the purpose in hand to supplement the field equations by a hypothetical relation between the distribution of the charge that constitutes the electron and its velocity. If this is done, it must be a relation which maintains its form in the correlated moving system that we have built up. Since it is a purely geometrical relation it must be a relation which is invariant under, or is a consequence of, the fundamental transformation (A). Hence Lorentz's assumption that the electron, which is naturally thought of as symmetrical round a centre when it is at rest, is, when in motion, of spheroidal shape, being obtained from the spherical stationary electron by the simple application to it of the FitzGerald contraction. If this is not assumed, the correlation between the stationary and the moving system is not perfect, and we have no reason, therefore, for thinking that the desired contraction of the whole body will take place.

Thus the whole explanation reduces to the formation of a conception of the electron, which really embodies the fact that we are trying to explain. The origin of the contraction is left in obscurity. The conception is only arrived at by assuming the result of the experiment to be, not an accident arising out of the particular circumstances of the case, but inherent in the constitution of matter down to its most minute elements. In other words, *the formation of this conception of the electron is itself a direct application of the Principle of Relativity.*

#### 47. THE EXPERIMENTS OF RAYLEIGH AND BRACE.

If we adopt the conception of the electron that has been described there is no need to go any further into the details of any phenomena which are explicable in terms of the motions of electrons within a body in order to be sure that we shall not be able to obtain a positive effect due to the motion of bodies through the æther. For as we have already remarked more than once, the adoption of a definite relation between the velocity and the configuration of the electron, so that the latter is determined uniquely by the former, makes the scheme

of the laws of the motions of the electrons complete, and the whole scheme is now a relative one; the æther is completely concealed.

It may, however, help to make the consequences of the conception clearer if we consider specially some of those results of it; especially some which directly affect certain experiments which were devised to detect effects which, it was thought, would follow from the contraction hypothesis.

It occurred to Lord Rayleigh that a shrinkage of a transparent body in one direction would cause it to cease to be isotropic, and that therefore there might arise a double refraction of a beam of light traversing it in a direction across the line of the motion. This was demurred to by Sir Joseph Larmor when the paper was read in 1902 at the British Association meeting on the ground of an argument similar to that given above. But the experiment was carried out again with extreme refinement by Brace in 1906, the only result being a complete fulfilment of the prediction of its failure.

The mechanism of refraction as ordinarily conceived is that the light waves travelling through the body set in vibration the electrons which go to make up the atoms, and their inertia, number, and distribution determine the amount of the refraction. If these are not isotropically disposed the body will produce a different refraction for different directions of the components of the electric force in the wave, and so a double refraction will arise. Now, Lorentz<sup>1</sup> was able to show that, starting from the conception of the electron which he had developed, the effective inertia of an electron is not the same for all directions. On allowing for this it appears that the inequality in the inertia is exactly of that amount which is required to neutralize the inequality in distribution produced by the contraction of the body. This, of course, in the light of what we have said above, is exactly what we should expect.

#### 48. THE EXPERIMENT OF TROUTON AND RANKINE.

In spite of the results of earlier experiments and of the theory of Lorentz and Larmor it still seemed to some that

<sup>1</sup> "Theory of Electrons," p. 217.



there might be need of putting the matter to a further test. Again reasoning from the FitzGerald contraction as producing a lack of isotropy in a body, Professor Trouton<sup>1</sup> suggested that there might be a difference of electrical conductivity in different directions in a conducting body carried along with the earth through the æther.

The conductivity of a body like its refracting power is interpreted in terms of the inertia and distribution of electrons which move within the body; and again, if the difference in apparent inertia for different directions is taken into account, the effect of the contraction on the distribution is annulled. The expectation of a positive effect is therefore unjustified, as we might predict on general grounds without any calculation of the inertia of the electron at all. The result of the experiment again verified this conclusion.

#### 49. THE EXPERIMENT OF TROUTON AND NOBLE.<sup>2</sup>

In this experiment, which arose out of a suggestion made by FitzGerald, a parallel plate condenser was suspended with the plates in a vertical plane and capable of turning about a vertical axis. An attempt was made to detect a tendency for the condenser to set itself in a definite direction when charged. This was expected to happen for the following reason: If a charged condenser is in motion through the æther, the electric force

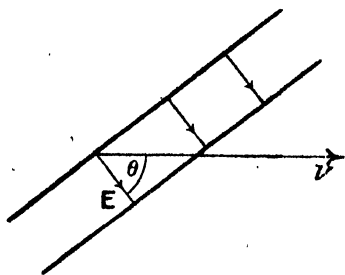


FIG. 5.

between the plates being  $E$ , then a magnetic force is set up in the region between the plates equal to  $Evsin\theta/c$ , to the first order in  $(v/c)$ , where  $v$  is the velocity of the condenser and  $\theta$  is the angle between the normal and the direction of motion. This magnetic field makes a contribution to the energy of the moving system proportional to  $E^2v^2sin^2\theta/c^2$ .

<sup>1</sup>"Proc. Roy. Soc." (A), 8 (1908), p. 428.

<sup>2</sup>"Proc. Roy. Soc.," 72 (1903), p. 132.

But in addition to this the electric force between the plates is altered by a quantity of the order  $E^2 v^2/c^2$ , and so there is a further change in the energy which is also of the order  $E^2 v^2/c^2$ . A calculation gives for the total change of energy the quantity

$$W = 4\pi CV^2(\sin^2 \theta - \frac{1}{2})v^2/c^2$$

where  $C$  is the capacity of the condenser and  $V$  the potential difference. If now the condenser is rotated so as to alter the angle  $\theta$  without altering the charge, work will have to be done in changing the energy; this indicates that there is a couple acting on the condenser of magnitude  $-\frac{dW}{d\theta}$

that is,  $-8\pi CV^2 \sin\theta \cos\theta v^2/c^2$ . This couple vanishes if  $\theta$  is equal either to 0 or to  $\frac{1}{2}\pi$ ; that is if the condenser is either normal or parallel to the direction of motion, the stable position being that in which the energy is a minimum, that is, when the condenser is normal to the velocity.

The method of observation was to reverse the charge of the condenser at regular intervals of time equal to half the time of the natural period of swing of the condenser. By this means it was hoped that the cumulative effect of the alternating couple would gradually set up a measureable swing in the system. The experiment was tried at different times of day so as to ensure that by no accident the actual velocity of the earth relative to the æther was normal to the plane of the condenser; but no effect could be detected in any case.

If the reasoning whereby we were led to anticipate this result is correct, the expression for the energy, which is obtained after allowing for the FitzGerald contraction, leads to a disagreement between the result of experiment and the conclusions drawn from applying the ordinary mechanical principles. This brings us to the point that was overlooked by the designers of the experiment. In experiments which take account of the second order in  $v/c$ , the Newtonian mechanics require correction. This is clear from the case of the electron that has already been referred to. The apparent inertia being different in different directions, the force on the electron is not in the direction of its acceleration. It will clearly be a matter of some difficulty to derive from a constitutive theory

of matter as composed of electrons of Lorentz's type the general way in which the dynamics of a rigid body are to be modified. The only feasible way of finding the kind of correction required is to start from the hypothesis of relativity in the new sense, not in the old dynamical sense, and to see what is the simplest modification that can be made in accordance with it, just as Lorentz's conception of the electron is the simplest type of nucleus that can be devised to accord with it. Some indication of the solution of this problem will be given in a later chapter, where a fuller discussion of the experiment will be found. (See p. 81.)

## CHAPTER V.

### MECHANICS AND THE PRINCIPLE OF RELATIVITY.

#### 50. THE TRANSITION FROM ELECTRICAL TO MECHANICAL THEORY.

IN making the transition from the purely electrical theory that we have been chiefly considering in the last two chapters to mechanical theory, it is necessary first to state clearly what is the connecting-link.

We might follow the method of the critical school which treats of force as a mere concept, the formation of which is possible on account of the physical laws of conservation of mass and momentum. This school of thought begins with the hypothesis, based on experience, that the accelerations of two particles which act on one another are in a constant ratio. This is, in effect, Newton's law of equal and opposite action and reaction, combined with the conception of a mass-ratio for two particles which is constant and independent of their velocities and of any other circumstances. In order to adapt this hypothesis to the requirements of the Principle of Relativity, we should have to modify it in such a way that the kinematic law which we put in its place has a form which remains the same when we change by a Lorentz transformation from one set of space-time co-ordinates to another. Having done this, we might define the force acting on a particle in a convenient way, and hence obtain the manner in which we conceive the measure or direction of a force to change when we change the frame of reference. Then subsequently we might bring the force, as so defined, into relation with the electrical field

by seeking for a quantity in terms of the electro-magnetic vectors which is subject to the same transformation as that which has been suggested on kinetic grounds. For any equation which embodies a relation actually subsisting in physical phenomena must, according to the Principle of Relativity, have a form which is independent of the particular frame of reference employed.

It may be as well to point out by an illustration that some modification is necessary.

The addition equation of velocities which is fundamental in the relativity kinematics renders the Newtonian form of the equation of the conservation of momentum no longer possible. If we consider two particles moving in a straight line under no forces save their own mutual action, we have the equation

$$m_1 u_1 + m_2 u_2 = \text{constant.}$$

If we change the frame of reference, by the use of the equation (a), p. 39, we have in place of this

$$m_1 \frac{u_1' + v}{1 + u_1' v / c^2} + m_2 \frac{u_2' + v}{1 + u_2' v / c^2} = \text{constant,}$$

which is of an entirely different form.

It is not easy, without further analytical assistance, to see in what way the original equation is to be modified in order to obtain a form which is the same in both frames of reference. Minkowski devised an analytical method of great power and elegance by which it can be done very quickly. To this we shall come in the next chapter. We turn, for the present, to a means of bridging the gap between electrical and mechanical theory, which is nearer to that which has been usual in electrodynamics, and which is followed by Lorentz, who makes the transition by introducing the expression for the 'mechanical force per unit charge,'  $\mathbf{e} + [\mathbf{uh}]/c$ .

## 51. APPLICATION OF THE LORENTZ TRANSFORMATION; MOTION OF A CHARGED PARTICLE.

We adopt this definition and proceed to consider the motion of a charged particle. In order to do this it is sufficient

to consider a transformation between two systems, in one of which the particle is instantaneously at rest. We shall assume that in this system the acceleration of the particle is given by

$$q \mathbf{e} = m \mathbf{f},$$

$q$  being the charge on the particle,  $\mathbf{f}$  the acceleration, and  $m$  being a definite constant. That is, we assume that for a particle whose velocity may be neglected, the motion is given by an equation of the ordinary Newtonian form,  $m$  being the 'apparent mass' when the velocity is zero.

Now, if we take the addition equations

$$u'_x = \frac{u_x + v}{1 + u_x v/c^2}, \quad u'_y = \frac{u_y}{\beta(1 + u_x v/c^2)}, \quad u'_z = \frac{u_z}{\beta(1 + u_x v/c^2)},$$

and assume that  $\mathbf{u}$  is very small, these give us to the first order in  $\mathbf{u}$

$$u'_x = v + u_x(1 - v^2/c^2), \quad u'_y = u_y/\beta, \quad u'_z = u_z/\beta.$$

Now, let us suppose that the velocity  $\mathbf{u}$  is the velocity communicated to the particle from rest in a short time  $\delta t$ ; the corresponding interval of time in the other frame of reference is to be obtained from the equation

$$\delta t' = \beta(\delta t + v \delta x/c^2)$$

in which we have to put  $\delta x$  equal to zero since the velocity of the particle is momentarily zero. Thus

$$\delta t' = \beta \delta t.$$

Hence, dividing the increments of velocity by the corresponding intervals of time, we have for the relation between the expression for the accelerations in the two systems

$$f'_x = f_x/\beta^3, \quad f'_y = f_y/\beta^2, \quad f'_z = f_z/\beta^2.$$

But we have also from the transformation of the electric and magnetic intensities,

$$e'_x = e_x, \quad \beta\left(e_y + \frac{v}{c} h'_z\right) = e_y, \quad \beta\left(e'_z - \frac{v}{c} h'_y\right) = e_z.$$

Thus the equations of motion become

$$q e'_x = \beta^3 m f'_x,$$

$$q\beta\left(e_y' + \frac{v}{c}h_z'\right) = \beta^2mf_z,$$

$$q\beta\left(e_z' - \frac{v}{c}h_y'\right) = \beta^2mf_x';$$

or, introducing the 'force' vector  $\mathbf{F}' = (\mathbf{e}' + [\mathbf{u}\mathbf{h}]/c)$ , since  $\mathbf{u}' = (v, 0, 0)$  we may write these equations

$$qF_x' = \beta^2mf_x' \quad qF_y' = \beta mf_y', \quad qF_z' = \beta mf_z'.$$

These are the equations which are satisfied by the motion of the particle relative to the system in which the particle has a velocity  $v$  parallel to the axis of  $x$ . They are commonly interpreted by saying that the particle has a 'longitudinal mass' equal to  $\beta^3m$ , and a 'transverse mass' equal to  $\beta m$ ; the word 'mass' being understood as the 'force per unit of acceleration produced,' and the 'force per unit charge' being defined to be  $(\mathbf{e} + [\mathbf{u}\mathbf{h}]/c)$ .

As far as the dynamics of the particle are concerned, the above equations rest on the assumptions, (i) that observations of the acceleration of a charged particle by an electro-magnetic field will not furnish us with a means of distinguishing between the different systems of co-ordinates for which the laws of purely electrical phenomena preserve their form; and (ii) that if the velocity of a particle relative to the frame of reference is sufficiently small, the law of its motion reduces to the Newtonian form, the form which has been adopted as the result of the observation of the motions of bodies whose velocities relative to the earth are all very small compared with the velocity of light.

## 52. THE EXPERIMENTS ON THE DEVIATION OF THE $\beta$ -RAYS BY AN ELECTRO-MAGNETIC FIELD.

We may now proceed to consider what light is thrown upon the equations and assumptions that have been written down above by the well-known experiments of Kaufmann and others on the inertia of the negative electrons. In the first form of these experiments a stream of the  $\beta$ -rays from a small piece of a radio-active substance passes through a fine hole in a screen and impinges on a photographic plate. Electric and

magnetic fields are then set up in a common direction at right angles to the stream; it is then found that, instead of meeting the screen in a single small spot, the stream produces on the plate a curved trace.

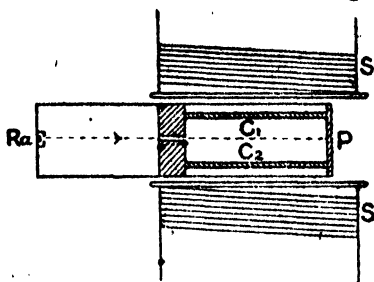


FIG. 6.— $C_1, C_2$ , condenser;  $P$ , photographic plate;  $S, S$ , electro-magnet;  $Ra$ , radium fluoride.

The interpretation of this is as follows. Supposing the rays to consist of a stream of small charged particles, they will be deflected by the electric and magnetic fields. Let us take the axis of  $x$  to be in the direction of motion, and the axis of  $y$  to be in the common direction of the fields; let the accelerations in the directions of the axes of  $y$  and  $z$  be  $f_y$  and  $f_z$ , the velocity along the axis of  $x$  being  $v$ ; then if  $l$  is the distance traversed through the field, and  $t$  is the time of transit,

$$l = vt;$$

and the deviations in the directions of the axes of  $y$  and  $z$  are

$$y = \frac{1}{2}f_y t^2 = \frac{1}{2}f_y l^2 / v^2,$$

$$z = \frac{1}{2}f_z t^2 = \frac{1}{2}f_z l^2 / v^2.$$

The measurement of the deviations  $y$  and  $z$  on the photographic plate thus gives the values of  $f_y/v^2$  and  $f_z/v^2$ .

Let  $E, H$  be the respective intensities of the electric and magnetic fields. Then, if the equations obtained above are a correct representation of the facts, we have, since with the given arrangement the components of  $\mathbf{F}$  are  $(0, E, vH/c)$ ,

$$qE = \beta m f_y, \quad qvH/c = \beta m f_z,$$

from which we have

$$\frac{vH}{cE} = \frac{f_y}{f_z} = \frac{y}{z}.$$

Thus, corresponding to any particular point on the trace of the ray on the plate, we can obtain an estimate of the velocity of the particles which strike the plate at that point, the strengths of the fields being supposed to be known. We are then able



to test the correctness of the assumptions that led to the above equations by means of either equation. The first gives us

$$\frac{q}{m} = \frac{\beta f_y}{E},$$

and each factor on the right-hand side is now a quantity of which we have an estimate. If it be true that all the particles forming the stream are identical in nature, then the values of  $q/m$  obtained from all the different points of the trace of the rays on the plate should be the same.

### 53. THE METHOD OF THE CROSSED FIELDS.

An improvement on this method was introduced by Bestelmeyer. It is now known as the method of 'crossed fields'.

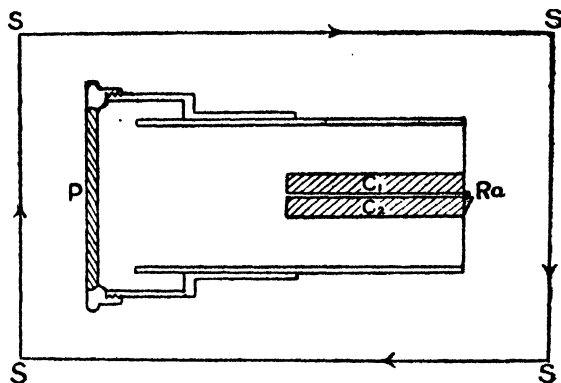


FIG. 7.

The figure shows the essential parts of the apparatus as used by Bucherer and Wolz.  $C_1$  and  $C_2$  are the plates of a condenser, consisting of two parallel optically-plane plates of glass silvered on the inner side, and kept at a fixed very small distance (about 25 mm.) by pieces of a thin sheet of quartz.  $Ra$  is a grain of radium fluoride from which the  $\beta$ -rays can pass between the plates of the condenser to strike the photographic plate  $P$ , which is at a measurable and adjustable distance from the condenser. The whole of this part of the apparatus is placed transversely inside a long solenoid, of which the rectangular section SSSS is shown in the figure. The electric field between the plates of the condenser and the

magnetic field due to the solenoid are now at right angles, and the  $\beta$ -particles move across them both. As they are emitted in all directions by the radium fluoride, they cross the magnetic field at different angles with its direction. If we consider one which moves with velocity  $v$  at an angle  $\alpha$  with the magnetic field  $H$ , it is subject to a force  $qvH \sin \alpha/c$  normally to the plates of the condenser, while, at the same time, it is subject to a force  $qE$  due to the electric field  $E$ . These two forces will balance one another if  $vH \sin \alpha = cE$ , and, if this is so, the electron will move in a straight line and so be able to pass right through the narrow gap between the plates of the condenser. (The gap between the condenser plates being so narrow compared with their length, it may be taken that the electrons emerging at a given angle  $\alpha$  have a very definite velocity given by this equation.) On emerging they will be subject to the magnetic field only, and will therefore be deflected from a straight path.

According to the above theory there should be an acceleration perpendicular to  $v$  and to  $H$  of magnitude  $qvH \sin \alpha/mc\beta$ , and the distance to be traversed is  $d/\sin \alpha$ ,  $d$  being the distance from the condenser to the plate.

Thus to a first approximation the deflection  $z$  will be given by

$$\begin{aligned} z &= \frac{1}{2}(qvH \sin \alpha/mc\beta) (d/v \sin \alpha)^2 \\ &= \frac{1}{2} \frac{qd^2 H^2}{m\beta c^2 E}. \end{aligned}$$

In this expression for  $z$  every factor is independent of  $\alpha$  except  $\beta^{-1}$  which is equal  $(1 - E^2/H^2 \sin^2 \alpha)^{\frac{1}{2}}$ .

Thus the electrons will meet the plate in a curved line with a maximum deflection for the rays incident normally, that is, for  $\alpha = \frac{1}{2}\pi$ .

The measurements in the third part of the table below were made by the application of this method, which has proved capable of much more accuracy than the original method of Kaufmann.

#### 54. TABLE OF RESULTS.

For the purpose of comparison of results obtained by different methods and for electrons of different velocities, it is

convenient to put together the values of the ratio  $q/m$ ,  $m$  being, as above, the apparent mass for small velocities, the equations of motion being assumed to be those obtained above by the application of the Principle of Relativity. It will be seen that there is a remarkable agreement between the values obtained.

The first four lines in the table give the ratio  $q/m$  for comparatively slow rays, as the cathode rays, and the rays emitted by a glowing body. For these there is no question as to the dependence of inertia upon velocity.

Next is given the value which is predicted by the best measurements of the Zeeman effect, interpreted on the hypothesis that the spectral lines are due to the vibrations of an electron.

The last part of the table gives the values obtained from  $\beta$ -rays of velocity comparable with that of light.

	$v/c$	$q/m \times 10^{-7}$
J. Classen, "Verh. d. Deut. Phys. Ges.," 10 (1908), p. 700 .	*	1'773
J. Malassez, "Ann. d. Chim. et de Phys.," 23 (1911), p. 231 .	*	1'769
A. Bestelmeyer, "Ann. der Phys.," 35 (1911), p. 909 .	*	1'766
E. Alberti, "Ann. der Phys.," 39 (1912), p. 1133 .	*	1'766
Weiss and Cotton, "J. de Phys.," 6 (1907), p. 429 .	†	1'767
P. Gmelin, "Ann. der Phys.," 28 (1909), p. 1079 .		1'771
A. Bucherer, "Ann. der Phys." { 28 (1908), p. 513 .	{ 31	{ 1'752 <sup>1</sup>
{ 29 (1909), p. 1063 .	{ 69	{ 1'767
R. Wolz, "Ann. der Phys.," 30 (1909), p. 273 .	{ 5	{ 1'7676
C. Schaefer (G. Neumann), "Phys. Zeit.," xiv (1913), p. 1117	{ 7	{ 1'7672
	4 to 8	1'7676 <sup>2</sup>

In addition to the experiments of Wolz, Bucherer, and Neumann, reference should be made to those of Hupka.<sup>4</sup>

\* Negligible.

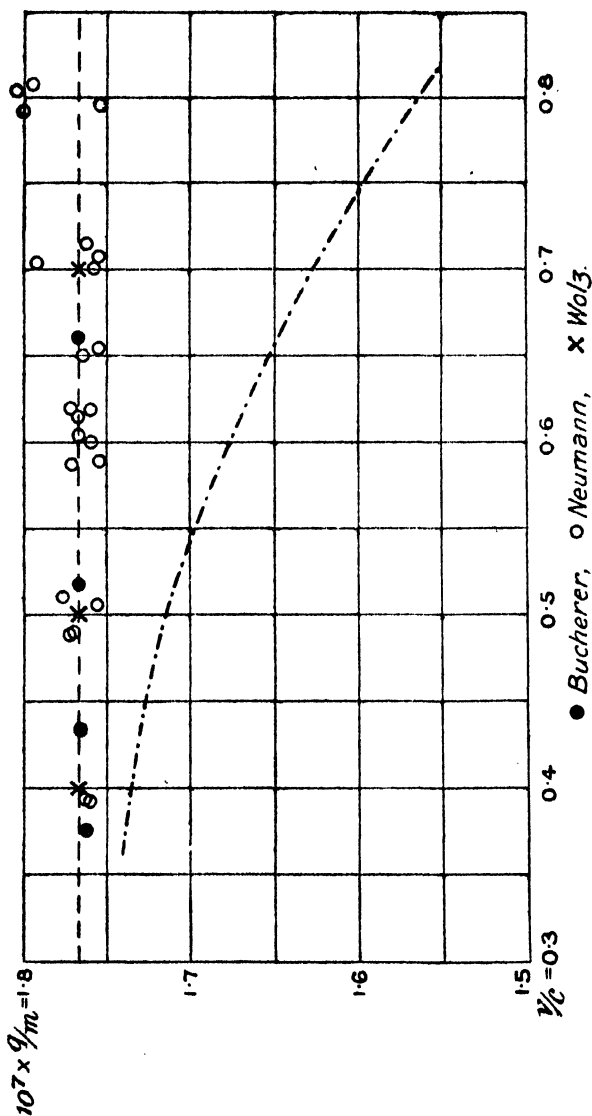
† Assumed to be negligible; experiments on Zeeman effect.

<sup>1</sup> Bucherer, in a later paper, points out a correction which should be made for the edge effect of the condenser. This would probably raise this result to the same value as the others. The following experiments of Wolz took particular note of this correction.

<sup>2</sup>  $\pm .0025$ .

<sup>3</sup> Of 26 results obtained, all lie between 1'75 and 1'8; all but four between 1'75 and 1'78. Possible error,  $\pm .0025$ . See Figure (8) below.

<sup>4</sup> "Ann. d. Phys.," 31 (1910), p. 169.



The curve — · — · — shows values of  $q_m$  calculated by the formula of Abraham

FIG. 8.

This author does not give an absolute value of  $q/m$ . He shows, however, that over a considerable range of velocities the formula of Lorentz gives a very constant value for  $q/m$ , e.g. in one set of Hupka's experiments in which the values of  $v/c$  varied between .346 and .501 the calculated values of  $q/m$  did not vary by more than .03 per cent.

### 55. THE SIGNIFICANCE OF THIS VERIFICATION.

The experiments which have just been discussed were, in the first instance, undertaken by Kaufmann merely with the intention of examining how far the inertia of an electron was due to the charge which it carried. It had been suggested by Sir J. J. Thomson as early as 1881 that the inertia of a body was increased by the presence of an electric charge on it. The result of Kaufmann's first experiments on the cathode rays was to show that, within the range of velocities there present, the whole inertia of the particles varied exactly in the same way as the electro-magnetic part of it should do, as calculated by Abraham from the hypothesis that the electron was a spherical distribution of charge.

Three years later, on the publication of the suggestion of the contracting electron to explain the results of the Michelson and Morley experiment, Kaufmann repeated his experiments with the  $\beta$ -rays in which higher velocities were accessible, with the object of distinguishing between the two theories. It cannot be said that the results obtained were entirely in favour of Lorentz's hypothesis; but more recent experiments, as the figures reproduced in the above tables show, were distinctly so. We may now say with certainty that the contraction hypothesis is not contradicted by experiment.

But to say that the results confirm the hypothesis 'that the whole of the inertia is attributable to purely electro-magnetic origin' is to forget the nature of the assumptions that are made in making the calculations.

The argument, as given by Lorentz, rests essentially on the hypothesis of the contracting electron. It was remarked above that to make this hypothesis is in reality only to shift the difficulty of explaining the facts back from the FitzGerald

hypothesis, which refers to concrete matter, to the same hypothesis about a purely conjectural element in the constitution of matter, and there it must be left. If we accept the crude picture of the electron as a spherical shell carrying a surface layer of charge, we must not ask how this shell is held together against the mutual repulsions of the parts of the charge which is carried; in fact, we hardly dare think of the charge on the electron as divisible, much less may we ask why the electron contracts, and whether there are 'forces' which cause it to do so. All that we can do is merely to assume that it is exactly so constituted as to fall in with the experimental facts.

There is then something over and above purely electro-magnetic theory required for the complete explanation of the facts observed. The simple comparison of the results of Abraham and Lorentz, which give entirely different values for the apparent inertia of an electron whose velocity is comparable with that of light, shows that the choice of this additional element to supplement the electro-magnetic laws is one which may seriously affect the results predicted by those laws. This is, for example, clearly shown by the lower curve in fig. 8, which exhibits side by side the values of  $q/m$  calculated from the experimental results of Neumann (i) by Lorentz's theory, (ii) by Abraham's theory. The latter theory differs only from that of Lorentz in the assumption that the electron is always a sphere, whatever velocity it is supposed to have.

Thus the agreement of experiment with the predictions of Lorentz's theory confirms the general correctness of his hypothesis. But it cannot be said to confirm the particular hypothesis of the contracting electron, which is only one of many that are consistent with the Principle of Relativity. In any case, it cannot be said that Lorentz's theory is a *purely* electro-magnetic one.

The results recorded above are therefore to be considered as confirming the general hypothesis of relativity as a principle that goes beyond the mere invariant form of the electro-magnetic equations.

## CHAPTER VI.

### MINKOWSKI'S FOUR-DIMENSION VECTORS.

#### 56. INTRODUCTORY.

THE position at which our discussion has now arrived is this. We saw in Chapter III. that, confining our attention to the facts of the propagation of light in free space, the possible methods of partitioning time and space were not independent of one another, that there was no unique sense to be attached to the phrase "simultaneity of events". This has been followed by a consideration of physical phenomena of a varied nature, with the result that we have arrived at the conclusion that the ambiguity in, and the interdependence of, the modes of measurement of time and space has, up to the present, not been removed.

Minkowski says of the situation so created: "*From henceforth, space by itself, and time by itself, are mere shadows, and only a kind of blend of the two exists in its own right*".<sup>1</sup>

All future discussion of the exact relations of the constitution of matter must be influenced by the general evidence for the validity of the hypothesis of relativity. If the principle is not universal it will sooner or later prove its own destruction by predicting results which are at variance with experiments. We must remember, however, that the principle is on a different plane from such "physical" laws as those embodied in Maxwell's equations or the law of gravitation. It is a general

<sup>1</sup> "Raum und Zeit," 1908. Reprinted in "Das Relativitäts prinzip," Leipzig, 1913.

mode of thought. We are not concerned with any particular quantitative law, but with the question as to whether all the quantitative laws of physics conform to a certain common criterion.

There is one such common criterion which is commonly accepted. We may call it the "vectorial" criterion, or the criterion that a mere change of orientation of a physical system in space will not alter its intrinsic properties. Space is said to be isotropic as far as it is merely a mode of ordering physical properties. The result is that, although we often employ a particular set of co-ordinate axes as a frame of reference, we recognize that the form of the equations which embody our physical laws must be one which is unaltered if we turn our axes into any other angular position. What is known as "vector analysis" is a notation and calculus which recognises this principle from the beginning, and which is adapted specially and only to the expression of such laws as conform to it.

The criterion of relativity is very analogous to this, and was, in fact, shown by Minkowski to be expressible as a generalization of it. An attempt will now be made to present the main idea introduced by Minkowski.

### 57. THE LORENTZ TRANSFORMATION AS A ROTATION OF AXES.

The change in the measures of space and time, which we have called a Lorentz transformation, may be written thus—

$$\begin{aligned}x' &= x \cos \theta - u \sin \theta, & y' &= y, \\u' &= x \sin \theta + u \cos \theta, & z' &= z,\end{aligned}$$

where,  $i$  standing as usual for  $\sqrt{(-1)}$ ,

$$u = ict, \quad u' = ict', \quad \cos \theta = \frac{c}{(c^2 - v^2)^{\frac{1}{2}}}, \quad \sin \theta = \frac{iv}{(c^2 - v^2)^{\frac{1}{2}}},$$

the last two quantities satisfying the equation

$$\cos^2 \theta + \sin^2 \theta = 1.$$

The analogous equations in three variables,

$$\begin{aligned}x' &= x \cos \theta - z \sin \theta, & y' &= y, \\z' &= x \sin \theta + z \cos \theta,\end{aligned}$$



are the equations which correspond to a simple rotation of the axes of co-ordinates through an angle  $\theta$  about the axis of  $y$ .

We may then think of the Lorentz change of variables as a rotation of the imaginary set of axes through the imaginary angle  $\theta$  in the four-dimension space in which the co-ordinates are  $(x, y, z, ict)$ . This is, of course, only a way of visualizing what are in effect algebraic processes involving only real quantities.

### 58. EXTENSION OF THE IDEA OF A "VECTOR".

Again, in three dimensions we speak of 'vector quantities,' as directed quantities which combine according to the parallelogram law.

If a certain vector is represented by components  $(X, Y, Z)$  along the axes of  $(x, y, z)$ , and by the components  $(X', Y', Z')$  when the axes of  $(x', y', z')$  are used, the relations connecting  $(X', Y', Z')$  with  $(X, Y, Z)$  are exactly the same as those connecting  $(x', y', z')$  with  $(x, y, z)$ . A vector quantity can be exactly represented by a line.

So Minkowski speaks of '**4-vectors**' as quantities with four components which when the frame of reference is changed are subject to exactly the same equations as are the space-time co-ordinates  $(x, y, z, u)$ . As a particular case, for the Lorentz transformation the 4-vector  $(X, Y, Z, U)$  is changed to  $(X', Y', Z', U')$ , according to the equations

$$X' = \beta(X - ivU), Y' = Y, Z' = Z, U' = \beta(U + ivX/c^2),$$

where the velocity  $v$  is arbitrary, and the direction of the axis of  $x$  is also an arbitrary direction in space. . .

These equations are equivalent to

$$X' = \beta(X + vT), Y' = Y, Z' = Z, T' = \beta(T + vX/c^2),$$

where  $U = ict$ , and  $U' = ict'$ .

### 59. EXAMPLES OF 4-VECTORS.

Let two point-instants be in a certain frame of reference  $(x_1, y_1, z_1, t_1)$   $(x_2, y_2, z_2, t_2)$ . Since  $(x_1, y_1, z_1, ict_1)$ ,  $(x_2, y_2, z_2, ict_2)$  are 4-vectors, so is their difference

$$(x_2 - x_1, y_2 - y_1, z_2 - z_1, ict_2 - ict_1).$$

If we write for the differences  $(\delta x, \delta y, \delta z, ic\delta t)$ , remembering the fundamental property that

$$\delta x^2 + \delta y^2 + \delta z^2 - c^2 \delta t^2$$

is an invariant (p. 40), it follows that

$$(\delta x, \delta y, \delta z, ic\delta t)/\kappa \delta t$$

is a 4-vector; that is, for a moving point

$$\mathbf{u} = \kappa(u_x, u_y, u_z, ic) \text{ is a 4-vector}$$

where

$$\kappa = \{1 - (u_x^2 + u_y^2 + u_z^2)/c^2\}^{-\frac{1}{2}}.$$

More generally,

if

$$\mathbf{X} = (X, Y, Z, U) \text{ is any 4-vector}$$

then

$$\dot{\mathbf{X}} = \kappa \frac{\partial}{\partial t} (X, Y, Z, U) \text{ is a 4-vector.}$$

where  $X, Y, Z, U$  are supposed to be any functions of the space-time co-ordinates  $(x, y, z, t)$  of a certain moving point.

The following may be stated to be 4-vectors without proof:—

$$\mathbf{e} = \{\mathbf{e} + [\mathbf{u}\mathbf{h}]/c, i(\mathbf{u}\mathbf{e})/c\}$$

$$\mathbf{h} = \{\mathbf{h} - [\mathbf{u}\mathbf{e}]/c, i(\mathbf{u}\mathbf{h})/c\},$$

where the symbols have the same meanings as in the fundamental equations, and the first three components of the 4-vector are the components of the ordinary vectors specified.

## 60. SPACE AND TIME AS TWO ASPECTS OF A UNITY.

We now see what Minkowski had in mind when he spoke of "only a blend of space and time existing in its own right". The 4-vector is thought of as a single entity, just as ordinarily we think of a force as a single entity. The *components* of a force do not exist by themselves. They are only convenient means of specifying the *force* which is one and definite. Just in the same way space and time co-ordinates are to Minkowski only particular and complementary aspects of a single fact or occurrence.

Analytically Minkowski transports himself to a space of four dimensions in which the distinction between space and time vanishes. In this four-dimensional region, the whole of space and time is portrayed in one construct. The motion of a moving point through all time is represented by a single curve, the points on the curve being ordered to correspond with the succession of events in time, but the interpretation of the curve as representing an ordinary motion is not unique; it depends upon the choice of the direction in the four-dimensional region which is chosen to be the time axis.

Thus three-dimensional kinematics becomes four-dimensional geometry. This relation extends further, it reaches into the domain of mechanical quantities. Three-dimensional dynamics can be interpreted as a four-dimensional statics, an exact generalization of three-dimensional statics. An example of this will be given shortly.

#### 61. A SECOND KIND OF FOUR-DIMENSIONAL VECTOR.

Before passing on to this, however, it will be well in passing to indicate the way in which Minkowski, thinking in terms of this four-dimension space, pictures the significance of the relativity of the electro-magnetic vectors  $\mathbf{e}$  and  $\mathbf{h}$ .

In ordinary three-dimension space an element of a straight line—i.e. a one-dimensional region—and an element of area—i.e. a two-dimensional region—may be equally well represented as vectors. But in four dimensions, of the elements of regions of one, two, and three dimensions, only the first and third are capable of representation as what we have called 4-vectors. A two-dimensional element will have to be represented by six components, corresponding to the number of ways of choosing two out of the four co-ordinates of a point.

Now Minkowski was able to show that the transformations (B, p. 48), affecting  $(e_x, e_y, e_z)$  ( $ih_x, ih_y, ih_z$ ) are exactly of the same form as those which affect the six components of a two-dimensional region in four-dimension space when the fundamental geometrical transformation of  $(x, y, z, t)$  is carried out. Hence he introduces a new conception; he defines a '**6-vector**' to be a quantity with six components which transforms in exactly this way. Then the validity of the transformation (B) is contained in the phrase

$$\mathbf{F} = (\mathbf{e}, i\mathbf{h}) \text{ is a 6-vector.}$$

One property which follows from this, and which may be deduced immediately from (B), is that

$$\mathbf{F}^2 = \mathbf{e}^2 - \mathbf{h}^2 \text{ is invariant.}$$

We might here extend the remark of Minkowski quoted above (p. 72), and say that "*from this point the electric intensity and the magnetic intensity apart from each other have no signi-*

ficance, but that only a single quantity compounded of the two exists in its "own right".

## 62. THE PRINCIPLE OF LEAST ACTION A GENERALIZATION OF THAT OF LEAST POTENTIAL ENERGY.

It is now possible to give a striking example of the way in which these conceptions give to dynamical equations a symmetrical form which is an exact generalization of the ordinary vectorial form of statical equations.

In electrostatics the 'total energy' of a field may be written

$$\frac{1}{2} \int e^2 dV,$$

where  $dV$  is an element of volume<sup>1</sup>; and there is a well-known theorem that *the electro-static field is such as to render this total energy a minimum subject only to the restrictions imposed on it by a given distribution of charge.* The generalization of this is:—

*If  $d\mathcal{V}$  represent an element of the four-dimensional space, and  $\int \mathcal{F}^2 d\mathcal{V}$  be called the 'total action,' then the electro-dynamic field is such as to make this total action a minimum, subject only to the restrictions implied by a given distribution of moving charges.*

This is only the expression in the language of four vectors of the known theorem that the equations of the electro-magnetic field are such as to make the time integral of the difference of the magnetic and electric energies a minimum. In ordinary notation this integral is

$$\frac{1}{2} \iiint (\mathbf{e}^2 - \mathbf{h}^2) dx dy dz dt.^2$$

In comparing these theorems we may note that whereas in the former the 'total energy' is a quantity which is independent of the frame of reference, so in the second the 'total action' is a quantity that does not change when we alter the space-time system of measurement.<sup>3</sup> For it has been pointed out above that  $\mathcal{F}^2$  is invariant, and the element of integration  $d\mathcal{V}$  is unaltered by a turning round of the axes.

<sup>1</sup> Remembering that we are using Lorentz's units.

<sup>2</sup> See Larmor, "Æther and Matter," pp. 82 ff.

<sup>3</sup> In changing from the variables  $(x, y, z, t)$  to  $(x', y', z', t')$  we substitute for  $d\mathcal{V}$  the quantity  $\frac{\partial(x, y, z, t)}{\partial(x', y', z', t')} d\mathcal{V}'$ , and it is easily seen that  $\frac{\partial(x, y, z, t)}{\partial(x', y', z', t')}$  is unity.

## 63. THE NEW MECHANICS—'FORCE' AND 'WORK' AS TWO ASPECTS OF A SINGLE CONCEPT.

In the classical mechanics there have always been two strains of thought. The two aspects of 'force' as 'the time rate of change of momentum,' and as 'the space rate of change of energy,' have with different writers been given different degrees of prominence. Galileo developed the former, Huyghens the latter. In the light of four-dimensional vectors the two ideas become unified, and differ only as partial aspects of a greater concept, just as space and time are unified and become partial aspects of the whole extension of the universe.

The equations of motion for a body moving as a whole with velocity  $\mathbf{u}$  without rotation are

$$\frac{d\mathbf{g}}{dt} = \mathbf{k},$$

$\mathbf{g}$  being the momentum, and  $\mathbf{k}$  the moving force.

The energy equation is  $\frac{dw}{dt} = (\mathbf{k}\mathbf{u})$

$w$  being the energy, and  $\mathbf{u}$  the velocity.

In order to bring these equations into line with the Principle of Relativity, we have only to assume

(i) *that if  $\mathbf{g}$  is the momentum and  $w$  the energy then  $(\mathbf{g}, iw/c)$  is a 4-vector,  $\mathbf{G}$ ;*

(ii) *that if  $\mathbf{k}$  is the force acting on the particle and  $\mathbf{u}$  its velocity then*

*$(\mathbf{k}, i(\mathbf{k}\mathbf{u})/c)$  is a 4-vector,  $\mathbf{K}$ ;*

(iii) *that the ordinary equations of Newton are only an approximation to the more exact equations*

$$\mathbf{K} = \kappa \frac{d\mathbf{G}}{dt}.$$

We have seen that  $(\delta x, \delta y, \delta z, ic\delta t)$  being a 4-vector,  $\delta t/\kappa$  is an invariant.

Hence if  $\mathbf{f}$  is any 4-vector, and  $\delta\mathbf{f}$  is a small change in it,

$$\kappa \frac{\delta\mathbf{f}}{\delta t} \text{ or, in the limit, } \kappa \frac{d\mathbf{f}}{dt}$$

is another 4-vector.

Now, if  $(\mathbf{K}) = (\mathbf{k}, i(\mathbf{k}\mathbf{u})/c)$  be a 4-vector, the equation

$$\mathbf{K} = \kappa \frac{d\mathbf{G}}{dt}$$

is an equation of invariant form, that is, it expresses relations which are in accord with the Principle of Relativity.

The first three components of this equation give

$$\mathbf{k} = \kappa \frac{d\mathbf{g}}{dt}$$

which if  $(v/c)^2$  be neglected reduces to the Newtonian form,

$$\mathbf{k} = \frac{d\mathbf{g}}{dt};$$

while the fourth component is, to the same order,

$$(\mathbf{k}\mathbf{u}) = \frac{dw}{dt}$$

which is the usual equation of energy.

The distinction between the schools of Galileo and Huyghens, between 'force' and 'work,' is here lost, and we see them each as partial aspects of a greater whole.

#### 64. RELATION BETWEEN MOMENTUM AND ENERGY.

The stipulation that  $\mathbf{g}$  shall be a 4-vector requires us to modify the usual relations which connect the momentum and energy of a moving particle with its velocity. Let us make the assumption that *for a particle at rest the momentum is zero*. Then for that case

$$\mathbf{g}_0 = (0, 0, 0, iw_0/c)$$

where  $w_0$  is the energy of the stationary particle.

Now, for any particle whatever we may choose a frame of reference of which the origin has momentarily the same velocity as the particle. The extended velocity vector  $\mathbf{u} = \kappa(\mathbf{u}, ic)$  then becomes

$$\mathbf{u}_0 = (0, 0, 0, ic),$$

since in this frame of reference the particle is at rest.

The two vectors,  $\mathbf{g}_0$ ,  $\mathbf{u}_0$  are thus in the same direction, and we may write

$$\mathbf{g}_0 = \frac{w_0}{c^2} \mathbf{u}_0.$$

Now *the equality of two vectors is an invariant relation*; so that, whatever the frame of reference, we must have for the same particle

$$\mathbf{g} = \frac{w_0}{c^2} \mathbf{u}.$$

The last equation resolves into

$$\mathbf{g} = \kappa w_0 \mathbf{u} / c^2 = \frac{w_0 \mathbf{u}}{c^2 (1 - \mathbf{u}^2/c^2)^{\frac{1}{2}}},$$

and  $w = \kappa w_0 = \frac{w_0}{(1 - \mathbf{u}^2/c^2)^{\frac{1}{2}}},$

$\mathbf{g}$  and  $w$  being connected by the relation

$$\mathbf{g} = w \mathbf{u} / c^2.$$

Of these results we may note the following consequences if we neglect higher powers of  $\mathbf{u}/c$  than the second :—

(i)  $w = w_0 (1 + \frac{1}{2} \mathbf{u}^2/c^2) = w_0 + \frac{1}{2} m \mathbf{u}^2,$

(ii)  $\mathbf{g} = w_0 \mathbf{u} / c^2 = m \mathbf{u},$  where  $m = w_0/c^2.$

The quantity  $m$  here enters in exactly the same way as the Newtonian mass; thus we are led to contemplate a relation between the mass of a particle and its energy. That it is a possible relation appears from the following figures. If the particle were to give out energy, as in the case of radioactive bodies, the apparent mass should, according to this relation, diminish. But if we take the actual rate of loss of energy by a gram-atom (225 grms.) of radium in its disintegration, which is about  $10^{12}$  ergs per hour, and divide by  $c^2$ , the rate of diminution of the mass would be only about  $10^{-6}$  gr. per year, which is quite inappreciable by ordinary methods.

In the case of large velocities the concept of 'mass' becomes meaningless, but we still have the relation

$$\mathbf{g} = w \mathbf{u} / c^2$$

which indicates that the convection of energy by a moving body implies an amount of momentum equal to the product of energy and velocity divided by  $c^2$ .

Thus if a body moving uniformly is radiating energy, it is also losing momentum. The equation

$$\mathbf{k} = \beta \frac{d\mathbf{g}}{dt}$$

then indicates that there is a force acting on the body even though its velocity remains constant.

It has sometimes been suggested as a paradox, that whereas ordinary electro-magnetic theory clearly indicates a resistance

to the motion of a radiating body through the æther,<sup>1</sup> yet the Principle of Relativity must allow that a body radiating equally in all directions must be capable of remaining in uniform motion through the æther if subject to no external force.

The above analysis resolves the paradox; it admits the resistance, but finds it to be equal to the rate at which momentum is lost owing to the radiation of energy, even though the velocity be unchanged.

This proposition of the momentum or inertia of energy is one which, in the development of mechanics subject to the Principle of Relativity, is universal. A familiar instance is to be found in the mechanics of the free æther as ordinarily stated. Poynting, in an analysis of electro-magnetic energy, suggested that, at any point of free space, there is a flux of energy represented by the vector  $c[\mathbf{EH}]$ ; while later, Abraham showed that we may maintain the proposition of the conservation of momentum as between æther and electric charge if we allow of a distribution of momentum in the æther of intensity  $[\mathbf{EH}]/c$ . Here again the flux of energy is equal to the momentum multiplied by  $c^2$ , in exact agreement with the corresponding result for a material particle.

#### 65. THE TROUTON-NOBLE EXPERIMENT.<sup>2</sup>

There is an apparent contradiction, somewhat similar to that referred to in § 64, in the case of a radiating body in uniform motion, which arises in respect of the moment of the forces acting on a system in uniform motion. It can be shown that, in order to reconcile the mechanical principles of conservation of momentum and conservation of energy in the æther with the known laws of propagation of electro-magnetic disturbances, it is necessary to assign to the æther a velocity which is not in general in the direction of the momentum. In the same way it is possible so to modify mechanical equations of any deformable continuous system, using the same methods that have been indicated above for modifying the dynamics of a single particle, that they may be in accord with the Principle of Relativity.

<sup>1</sup> See Larmor, "International Congress of Mathematicians," Cambridge, 1912, p. 216.

<sup>2</sup> See p. 58.



In doing this two important results emerge. The first is that the relation referred to above between the flux of energy<sup>1</sup> and the momentum must be made universal. The second is that, in general, *the momentum per unit volume at any point of the medium need not be in the direction of the velocity of the medium at that point.*

In a case where this difference of direction arises it is easy to see that, for an element of volume moving with uniform velocity and with constant linear momentum, *the moment of momentum about a fixed point is not constant* as it is when the directions are the same. For in the figure let P, P' be the positions of the element at two consecutive instants, so

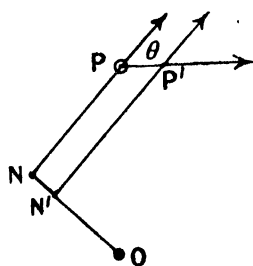


FIG. 9.

that  $PP' = v\delta t$ , and let the linear momentum be in the direction NP and of constant magnitude  $g$ .

Then it is clear from the figure that in the time  $\delta t$  the moment of the momentum about O diminishes by  $g \times NN'$ , that is, by  $gv \sin \theta \delta t$ , a quantity which only vanishes if  $g$  and  $v$  are in the same straight line.

Thus in the mechanics of the Principle of Relativity we shall not expect that the moment of momentum of a body moving uniformly relative to a given frame of reference will always remain constant. Exactly such a case arises in the conditions of the experiment of Trouton and Noble (see p. 58). Here a charged plane condenser is supposed to be placed obliquely to its direction of motion relative to a given frame of reference. The classical theory shows clearly that there must be a couple acting on it in a sense which would, according to ordinary dynamics, tend to turn it so as to be normal to the direction of motion. But if the dynamics of the Principle of Relativity are applied to this case we find that the couple is exactly equal to the rate of change of the moment of momentum which arises as above from the fact that the linear momentum, though constant, is in a different direction from the velocity. Thus the mechanical

<sup>1</sup> Flux of Energy = Momentum  $\times c^2$ .

conditions are completely fulfilled without any rotation of the condenser being set up.<sup>1</sup>

## 66. EXTENSION OF ANALOGY BETWEEN DYNAMICS AND STATICS.

In § 62 we have shown that the relation of the laws of the electro-dynamic field to the principle of least action can be looked upon as the extension to Minkowski's four-dimensional world of the relation of the laws of the electro-static field to the principle of least potential energy. In that statement, however, no account was taken of the mechanical constraints imposed upon the charge or current; the distribution of these was assumed to be given, and the variation of energy or current was subject to the given distribution.

In the electro-static case, if the charges or conductors carrying them are supposed to be moved, we know that the increment in the electro-static energy is equal to the virtual work of the non-electrical forces which maintain the equilibrium of the conductors against the electrical attractions and repulsions. Or again, if the non-electrical forces are conservative so that we may speak of a corresponding potential energy, the variation of the total potential energy in a small displacement from the equilibrium position is zero. We note, of course, that the total energy is a scalar quantity, and is independent of the choice of any particular system of reference.

In the dynamical case we have the corresponding theorem that if  $A_1$  represents the electro-dynamic action as defined in § 62, and if there be a non-electrical action  $A_2$ , arising from non-electrical inertia and forces, then if  $A = A_1 + A_2$  is called the 'total action,' then, for any slight variation of the currents and motion of the carriers of charge and current from the actual motion which takes place, the variation in  $A$  is zero.

This very general proposition must, as in the simpler case, be independent of the choice of a frame of reference; that is, if we maintain the relativity of all the phenomena, it must have an invariant form for any change of axes in Minkowski's four-dimensional space. Now we know, as in § 62, that  $A_1$ ,

<sup>1</sup> For a fuller discussion see the author's larger work (Camb. Univ. Press, 1914, p. 170).

the electro-dynamic part of the action, is an invariant. Hence the Principle of Relativity requires that if there is a non-electrical action  $A_2$ , it too must be an invariant.

70. One simple development may be set down as an indication of the application of the above result.

If  $\mathbf{k} = (k_x, k_y, k_z)$  is a typical force acting at a point  $(x, y, z)$  at time  $t$ , and we consider a slightly varied motion of the system in which the co-ordinates of the same point are  $(x + \delta x, y + \delta y, z + \delta z)$  at time  $(t + \delta t)$ , then the variation in the action arising from  $\mathbf{k}$  is known to be

$$k_x \delta x + k_y \delta y + k_z \delta z - (\mathbf{k}\mathbf{u})\delta t,$$

where  $\mathbf{u}$  is the velocity of the point in question (see e.g. Routh, "Rigid Dynamics," Vol. II, p. 280).

In Minkowski's notation this can be written  $\mathbf{k}\delta\mathbf{r}$  where  $\mathbf{k} = (\mathbf{k}, i(\mathbf{k}\mathbf{u})/c)$  and  $\delta\mathbf{r} = (\delta x, \delta y, \delta z, ic\delta t)$ . The invariance of this quantity,  $\delta\mathbf{r}$  being an arbitrary 4-vector, requires that  $\mathbf{k}$  shall also be a 4-vector.

## APPENDIX I.—THE FUNDAMENTAL EQUATIONS IN SYMMETRICAL FORM.

It may be worth while to set down the equations of the field as Minkowski arranges them. For fuller details see the author's larger work, Chapter IX.

Writing

$$(F_{yz}, F_{zx}, F_{xy}, F_{xu}, F_{yu}, F_{zu}) = (-H_x, -H_y, -H_z, iE_x, iE_y, iE_z),$$

and

$$F_{yz} = -F_{zy}, \text{ etc. ;}$$

also

$$(S_x, S_y, S_z, S_u) = \rho(u_x, u_y, u_z, ic),$$

and

$$u = ic t,$$

the equations of the field become

$$\left. \begin{aligned} \frac{\partial F_{yz}}{\partial y} + \frac{\partial F_{zx}}{\partial z} + \frac{\partial F_{ux}}{\partial u} &= S_x \\ \frac{\partial F_{xy}}{\partial x} + \frac{\partial F_{zy}}{\partial z} + \frac{\partial F_{uy}}{\partial u} &= S_y \\ \frac{\partial F_{xz}}{\partial x} + \frac{\partial F_{yz}}{\partial y} + \frac{\partial F_{uz}}{\partial u} &= S_z \\ \frac{\partial F_{xu}}{\partial x} + \frac{\partial F_{yu}}{\partial y} + \frac{\partial F_{zu}}{\partial u} &= S_u \end{aligned} \right\}$$

$$\left. \begin{aligned} \frac{\partial F_{uz}}{\partial y} + \frac{\partial F_{yu}}{\partial z} + \frac{\partial F_{xy}}{\partial u} &= 0 \\ \frac{\partial F_{xu}}{\partial x} + \frac{\partial F_{xz}}{\partial z} + \frac{\partial F_{xy}}{\partial u} &= 0 \\ \frac{\partial F_{uy}}{\partial x} + \frac{\partial F_{xu}}{\partial y} + \frac{\partial F_{yz}}{\partial u} &= 0 \\ \frac{\partial F_{yz}}{\partial x} + \frac{\partial F_{xz}}{\partial y} + \frac{\partial F_{xy}}{\partial z} &= 0 \end{aligned} \right\}$$

The symmetrical form of the equations is now obvious. They may be easily shown to have an invariant form, and may be looked upon as the generalization of the vectorial equations of the electro-static field

$$\operatorname{div} \mathbf{e} = \rho$$

$$\operatorname{curl} \mathbf{e} = \mathbf{o}.$$

$(S_x, S_y, S_z, S_u)$  is easily seen to be a 4-vector by means of the equations (a) and (C), p. 48.

## PART II.

# THE GENERAL PRINCIPLE OF RELATIVITY

## CHAPTER VII.

### THE GENERAL THEORY.

#### 67. THE NEED FOR MORE RADICAL RE-CONSTRUCTION.

WE are now in a position to appreciate the way in which a number of different difficulties united in calling for the comprehensive reconstruction effected by Einstein under the name of the General Principle of Relativity. The Special Principle was unsatisfying in two ways.

On the one hand, while demanding a drastic revision of our ordinary ways of thinking about space and time, it was by no means thorough-going enough for the philosopher. The old problem of absolute acceleration was left exactly at the same point at which Newton had left it.

On the other hand, the new principle had not been able to go any further than the electron theory of matter, of which it was a natural complement, in interpreting the nature of gravitation.<sup>1</sup> The electro-magnetic theory seemed fairly well able to account for those properties in which various kinds of matter differ from one another, as, for example, thermal and electrical conductivity, optical dispersion, magnetizability, as well as being a clear basis of chemical relations. But of the two universal and common properties of matter, inertia and gravitation, the latter remained obstinately aloof. At the same time, however, the growing conviction that the Newtonian conception of mass as an

<sup>1</sup>For attempts to solve the problem of gravitation on the ground of the Special Principle of Relativity, see Poincaré, "Rend. del Circ. Mat. di Pal.," 21 (1906), p. 166; also de Sitter, "Monthly Notices of Roy Astr. Soc.," March, 1911, p. 388; also various papers by Nordstrom, and others, in "Phys. Zeitsch."

invariable measure of the inertia of a body was at best an approximation, began to raise the question as to whether the gravitational attraction between two bodies would not also in some way depend upon the frame of reference adopted. For the new theory, supported by the experiments on the mass of the electron (Chapter V.), showed that the mass of a body for the purposes of inertial phenomena was certainly a function of the velocity, and probably of the internal energy of the body. The question at once arose: *Is the gravitational mass of a body, the quantity which measures the power of one body to attract another, also a variable quantity? If so, is it always proportional to the inertia-coefficient?*

These two questions bear one upon the other at various points. For instance, if Newtonian dynamics is only approximately correct, then the partial relativity of dynamical relations is only approximate too. Further, if energy of all kinds is inseparable from momentum, if light has inertia, may it too be subject to gravitation? This would imply a possible deviation from the electro-magnetic equations of Maxwell out of which the early Principle of Relativity arose. That being so, the velocity of light might not in strictness conform to the hypothesis of a fixed and uniform velocity.

This, of course, would be expected by one who adopted the point of view that space and time are merely arbitrary modes of ordering phenomena, and that from an *a priori* point of view we can know as little about absolute rotation or acceleration as about absolute velocity.

But *once concede that the velocity of light is not an ultimate, universal, and invariable quantity, and the whole basis of the Special Principle of Relativity needs revision.* For the Lorentz-Einstein transformations were just selected as that group which preserved this one metric property of the universe intact. Minkowski's four dimensional vectors also depend for their validity on this special hypothesis.

#### 68. EÖTVÖS' EXPERIMENTS.

The first point to be considered was, that of the relation between the two universal properties of matter, inertia, and gravitational attraction. In the first instance, the question

asked was merely this: *Is the inertia mass strictly the same as, or proportional to, the gravitation mass?* Or more precisely, within what limit of accuracy can we definitely affirm on experimental grounds, that they are so?

Here the experiments of Eötvös have been quoted,<sup>1</sup> which purport to show that the ratio of the gravitational attraction of the earth upon different bodies to their inertia masses certainly does not vary by as much as the fraction  $5 \times 10^{-8}$ ; this in a comparison between specimens of such different bodies as brass, glass, antimonite, and cork; bringing a mass of air into the comparison also,  $10^{-5}$  is the limit of accuracy within which it is certain that there is no variation.

The principle of the experiments is as follows:—

Two masses of different substances are fastened to the ends of the beam of a delicate torsion balance. The beam is adjusted to hang in an east to west direction. It hangs in equilibrium under the attraction of the two gravitational attractions, and the centrifugal forces arising from the earth's rotation. The latter are proportional to the *dynamical* masses of the bodies, the former to what we have called their *gravitation* masses. If the dynamical masses and gravitational masses are equal, the resultant forces (or apparent weights of the masses) will be parallel, and so their resultant will be a single force, which, of course, must just be balanced by the tension in the suspending thread. If, however, the masses are not exactly proportional, the apparent weights will not be parallel, and so can only be combined into a single force together with a couple about its line of action, balanced respectively by the tension and the torsion in the thread. The magnitude of this torsional couple will be approximately  $w l \phi$ , where  $w$  is the weight of either mass,  $l$  its distance from the point of attachment of the thread, and  $\phi$  the angle between the apparent weights—an angle which will depend upon the inequality of the mass ratios.

Eötvös thus describes his experiment:—

“I attached separate bodies of about 30 grms. weight to

<sup>1</sup> “Mathematische und Naturwissenschaftliche Berichte aus Ungarn,” Bd. 8 (1891), p. 64.

the ends of a balance beam about 25 to 30 cms. long, suspended by a thin platinum wire in my torsion balance. After the beam had been placed in a position perpendicular to the meridian, I determined its position exactly by means of two mirrors, one fixed to it, and one fastened to the case of the instrument. I then turned the instrument, together with the case, through  $180^\circ$  so that the body which was originally at the east end of the beam now arrived at the west end. I then determined the position of the beam again relative to the instrument."

If a couple such as has been mentioned above existed, it would on turning the instrument round act in the reverse direction. Here Eötvös continues—"If the resultant weights of the bodies attached on either side pointed in different directions, a torsion in the suspending wire should ensue. But this did not occur in the case in which a brass sphere was constantly attached to the one side, and glass, cork or crystal antimony was attached to the other, and yet a deviation of 1-60,000th of a second in the direction of the gravitational force would have produced a torsion of one minute, and this would have been accurately observed."

The actual deviation of the pendulum from the true vertical due to centrifugal force is about 360 seconds, so that to the accuracy of 1 in  $2 \times 10^7$  the ratio of dynamical to gravitational mass is shown to be the same for the two bodies.

More recent experiments by Zeeman<sup>1</sup> on bodies with radioactive properties show the same result to a degree of accuracy varying between  $2 \times 10^{-7}$  and  $3 \times 10^{-8}$ .

## 69. GRAVITATION AND A BEAM OF LIGHT.

If we are led by these considerations to think that the internal energy of a body must be as much an essential factor in determining its susceptibility to gravitational influence as it is in determining its inertia, a natural sequel is to suppose that the purely non-material energy in a light wave, and the inertia which it has been shown to possess, will also in some way be subject to gravitation also. But a difficulty at once presents itself. In the case of material bodies, gravitational action always shows itself in accelerated motion. It has been pointed

<sup>1</sup> Zeeman, "Verh. d. k. Ak. van Wet.," 26 (1917), p. 451.



out (§ 67) however that, as one of the consequences of the special Principle of Relativity, we are no longer bound to think of the momentum of a body as being in the same direction as, or bearing a constant ratio to, its velocity. The velocity may remain constant, but a change in the internal energy or stress may yet produce a change in the momentum.

Thus we have to ask the question: *If there is a gravitational action upon light, will it be evidenced in a change of velocity, or simply in a change in the nature of the disturbance which is being propagated?*

To ask the question is to admit that either alternative may prove true. It is in following the consequences of the first possibility that we see the generalization of the Principle of Relativity open out.

The first step in the quest is to consider the consequences in the ordinary material phenomenon of weight. We come back to Galileo and his discovery that falling bodies have all the same acceleration. Does light then share in this acceleration?

Another way of putting the same question is this:—

“To an observer who was falling freely in a gravitational field, would a ray of light appear to be travelling uniformly in a straight line?”

This form of the question suggests a further and more general consideration. Would it be true that to such an observer all perceptible influence of gravitation on physical phenomena of every kind would be lost? Or, if the acceleration were only a part of that of free fall, would the gravitational influence on all kinds of phenomena be diminished in the same ratio?

If that were the case it would appear that the nature of gravitational influence is strictly comparable with the nature of those effects that arise from an acceleration of material systems. Gravitational forces would be of the same nature as the so-called centrifugal force apparent in the case of rotating systems, but dismissed by the Newtonian as fictitious.

## 70. ECONOMY OF THOUGHT.

The whole of the discussion of the special Principle of Relativity was guided by the main thought of preserving the

simplicity of the relations conceived to hold between the physical phenomena. No one ever denied that other frames of reference with any accelerations or rotations and arbitrary changes of co-ordinates might be adopted at pleasure by those who did not mind giving up that simplicity. But once it is granted that those simple relations cannot in all circumstances be maintained as strictly true, the case for a limitation of the allowable frames of reference breaks down. As long as Newtonian dynamics were the last word in physical theory, it was of no avail for the philosopher to say that the mind could not conceive of absolute rotation. The scientist could only reply that nature certainly made a distinction between one frame of reference and another; that for some frames Newton's simple relations did hold, and for others they did not. But now the retort is invalid. The Newtonian equations are only approximately true in any case. The apparent simplicity arises because the frame of reference is chosen in such a relation to the phenomena in question that the remaining terms become inappreciable.

## 71. A FRESH START.

With this in mind we must turn back and see how far progress is possible along the lines of supposing that all preconceptions are given up as to the precise form of physical laws.

We are to remember that the use of space and time co-ordinates is simply a means to describing the relations that hold between physical phenomena; and that, as was said above (§§ 26 and 27), if we assume in advance a specific form for some physical law, we are in reality as much describing the relation of our system of measurement to the phenomena as describing the sequence of phenomena themselves. This will perhaps be made a little clearer by an illustration from two dimensions.

Two maps of a given part of the earth's surface, say a map on the Mercator projection, and a map on the stereographic, can easily be recognized, in spite of their difference in shape, as representing the same region. The ordering of places is distorted but not broken. If, on the map, the curves are

drawn which represent the shortest distance between given points (the great circles), the form of these curves will enable us to recognize what system of projection is being used. In other words the map embodies within it not only some information about the reality which it pictures; but, if the geodesics are drawn upon it, it gives also information about the mode of representation.

We are to think of the metrical description of physical phenomena as of a four-dimensional map of the history of the universe. The 'world lines,' as Einstein calls them, are the curves which in the map represent the history of identifiable particles. If we do not make any assumptions in advance as to the form of the world lines and their mutual relations, our mode of representation is quite arbitrary. By changing to another mode of representation the world lines may take quite different forms. The hypothesis of relativity assumes that no one form is more true or actual than another. They are all equally valid representations of reality.

It seems at first sight that, if we contemplate the distortion and stretching of the map to an arbitrary degree, there is no possibility of any metrical property of the universe being equally well represented on the map whatever form it may take. Yet an exact physical law must be a statement of equality which holds whatever method of expressing it is employed. The problem set to himself by Einstein and satisfactorily answered with the aid of pure mathematics already devised in other connections, was just that of finding mathematical relations which satisfied this criterion.

## 72. AN ESSENTIAL CONDITION OF MATHEMATICAL REPRESENTATION.

We must suppose that there is at least this much in common between all the possible pictures that may be formed of the universe and its history, that there is what we may call an exact correspondence of events. If something happens, a collision of two atoms, a tick of a clock, each observer will mark it as happening at a certain place and a certain time; and though they do not use the same methods of describing

the place and the instant, there will be in each of their several four-dimensional maps of all the happenings a definite point corresponding to and recording this event.

It must be a definite principle that there is what mathematicians call a one-to-one point-correspondence between all possible representations of physical phenomena.

An important application of this is in respect of what has been said above about the nature of experimental observations. To use again the example of the reading of a galvanometer, we merely observe a coincidence of a spot of light with a mark on a scale. This is what we have just called an event. In each several map of the universe the event is uniquely recorded as a single coincidence.

Suppose that  $(x_1, x_2, x_3, x_4)$  are the four variables which are used by one observer to define particular events. To any particular event corresponds a unique quartet of numbers as the corresponding values of these quantities, and vice versa to any quartet of numbers corresponds only one event. If another observer defines the event by another set of variables  $(x'_1, x'_2, x'_3, x'_4)$  the same is true of these.

Thus to a given set of values  $(x_1, x_2, x_3, x_4)$  corresponds a unique set  $(x'_1, \dots, x'_4)$  and vice versa.

### 73. THE COMPLETE RELATIVITY OF CO-ORDINATE SYSTEMS.

The special Principle of Relativity has taught us there is no distinction in character between the time and space co-ordinates. On the other hand, the metaphysician who has denied the validity of an absolute scale of time, or of space, has usually done it merely on the ground that, so long as the ordering in time and the ordering in space are not broken, a non-uniform stretching of the time variable or an arbitrary distortion in space may be contemplated.

Put mathematically the metaphysician (previous to the Special Principle of Relativity) was inclined to allow that—taking the suffix 4 to refer to the time variable we might allow that what one man calls a fixed point another may take to be moving in any manner he pleases—so that

$$x_1' = f_1(x_1, x_2, x_3, x_4)$$

$$x_2' = f_2(x_1, x_2, x_3, x_4)$$

$$x_3' = f_3(x_1, x_2, x_3, x_4)$$

$f_1, f_2, f_3$  being any functions whatever, while a non-uniform stretching of the time scale would be given by  $x_4' = f_4(x_4)$ .

With the assimilation of space and time variables proposed by Minkowski, the only general method of procedure seems to be to start by allowing  $x_1', x_2', x_3', x_4'$  each to be any arbitrary function of  $(x_1, x_2, x_3, x_4)$ .

We return then to the point at which we started (Chapter III, § 26) to construct all possible systems of space and time measurement, leaving aside the hypothesis of constant velocity, the only limiting hypothesis there employed.

#### 74. THE GENERAL QUADRATIC FORM.

That hypothesis was bound up with the invariant form of the quadratic differential expression

$$dx^2 + dy^2 + dz^2 - c^2 dt^2.$$

If  $x'$  is any function of  $x, y, z, t$ , say  $f(x, y, z, t)$

$$dx' = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial t} dt.$$

Thus on substituting expressions such as this for  $dx', dy', dz', dt'$  we find that

$$dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2$$

is identically equal to an expression of the form

$$a_{11}dx^2 + 2a_{12}dxdy + \dots$$

in which the coefficients  $a_{rs}$  are functions of  $x, y, z, t$ . The aggregate of these coefficients we shall call the tensor  $a_{rs}$  (see below § 79). With an *arbitrary* change of co-ordinates, the numerical values of the coefficients are not kept unaltered as they are under the Lorentz transformations; though the general form of the expression as a homogeneous quadratic differential form remains.

#### 75. SOME CONSEQUENCES OF A GENERAL CHANGE OF CO-ORDINATES.

Suppose now that  $x', y', z', t'$  are ordinary space-time co-ordinates for the representation of physical phenomena in a field in which the laws are of the classical form, in which, for

example, a wave of light is propagated uniformly with velocity  $c$ ; and in which there is no gravitation. In this picture, if we follow a light disturbance,

$$dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 = 0.$$

If the same phenomena are pictured by means of relations between  $x, y, z$ , and  $t$ , the same disturbance will be propagated according to the law

$$a_{11}dx^2 + 2a_{12}dxdy \dots + a_{44}dt^2 = 0.$$

If we seek for the velocity of propagation in a direction given by

$$dx = lds, dy = mds, dz = nds,$$

$ds$  being the actual distance travelled we have

$$ds^2(a_{11}l^2 + \dots + a_{33}n^2) + 2dsdt(a_{41}l + a_{42}m + a_{43}n) + a_{44}dt^2 = 0$$

which is a quadratic equation for the velocity  $ds/dt$ . For the moment the only interest in this equation is that the circumstances of the propagation depend upon the whole set of coefficients

$$\begin{vmatrix} a_{11} & \cdot & \cdot & a_{41} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{14} & \cdot & \cdot & a_{44} \end{vmatrix}$$

In other words, the *physical* properties of space as conceived by the observer using the co-ordinates  $x, y, z, t$  will depend upon this set of coefficients, though to the observer using the co-ordinates  $x', y', z', t'$ , space is homogenous and isotropic.

Further, if this latter observer notes a particle moving freely, that is, in terms of Newtonian laws, moving uniformly in a straight line, the observer who pictures its motion by means of  $x, y, z, t$  will picture it as moving neither uniformly nor in a straight line.

For example, if

$$x' = x \cos \omega t + y \sin \omega t \quad z' = z$$

and

$$y' = x \sin \omega t - y \cos \omega t, \quad t' = t,$$

a point at rest as mapped in the co-ordinates  $(x, y, z, t)$  will be describing a circle as mapped in the co-ordinates  $(x', y', z', t')$ , the angular velocity being  $\omega$ . Similarly a particle

moving uniformly in a straight line in the former co-ordinates would appear to be moving in a spiral round the origin in the latter.

Newton observing these phenomena would infer the existence of some influence exerting a 'centripetal force' on the moving particle. Einstein is aware of the fact that he is only observing the facts from a special point of view, and finds that by adopting another picture of the phenomena the necessity for postulating the force does not arise.

More generally, taking any arbitrary transformation of co-ordinates, and using quasi-geometrical language, the history of a moving particle is recorded by a curve drawn in the four-dimensional map of phenomena. By an arbitrary change of co-ordinates the map is stretched and distorted in an arbitrary manner, subject only to its not being split—otherwise one event in one representation would be split into two distinct events in the other. Newton, taking only one representation, infers from the straightness or otherwise of the history-curves (world lines) of particles their freedom to or subjection to external force. To Einstein 'force' is a purely relative term, relative, that is, to the particular picture he employs.

But we note here, in the passage from one picture to the other, that the same quantities which affect the unequal velocity of propagation of light affect the motion of the free particle. In fact, in the field of  $x', y', z', t'$ , the velocity of light is given by  $ds' = 0$  and the path of a free particle is such that  $\delta ds'$  is a minimum.

As pictured in the map with  $(x, y, z, t)$  co-ordinates the principle of exact correspondence shows that since numerically  $ds' = ds$  the velocity of light is determined by

$$ds = 0$$

and the path of a free particle by  $\delta ds$  taking a stationary value. Since  $ds^2 = a_{11}dx^2 + \dots$  the coefficients  $a_{11} \dots$  determine the course of both phenomena.

## 76. THE UNIVERSAL SIGNIFICANCE OF THE TENSOR $a_{rs}$ .

We now see why the tensor  $a_{rs}$  is associated by Einstein with the gravitational field. Gravitation is the name we have

given to the unseen influence which is evidenced by particles, otherwise free, exhibiting non-uniform motion. Granting the hypothesis of complete relativity we cannot say of a deviation or change of speed in the motion of a particle, whether it is due to the particular frame of reference used, or whether it arises from entirely external causes. We cannot distinguish between a 'real' gravitational field and a 'simulated' field. We can only deal with the actual motions perceived, and class the whole set of quantities characterizing those motions as defining the gravitational field, remembering, however, that the field is relative to the observer's particular system of measurement.

In the case of what we may call a completely simulated field, such as has been considered in the preceding section, we see that the whole set of quantities  $\alpha_{rs}$  is necessary to describe the field as apparent to any particular observer. We have seen that in this case there is a quadratic expression<sup>1</sup>

$$\alpha_{rs}dx_rdx_s$$

which for a given pair of adjacent events has the same value for all observers. We may say that the quantities  $\alpha_{rs}$  define the extent to which the space-time system used departs from isotropy and homogeneity in regard to its relation to physical events.

In the more general case of the actual universe it is difficult to imagine that the whole of what we have been in the habit of calling "gravitation" can be classed as a fictitious field of force. Nevertheless, if we are to grant the complete relativity of our space-time notions we must grant that we are not in a position to say to what extent it is fictitious in any given frame of measurement. The only plan of proceeding, therefore, seems to be to develop a scheme of relations more general than those which have been taken to hold in a purely fictitious field of force, but including those as a particular case.

All that has been assumed so far of that particular case is that when the fictitious field has been transformed right away the quadratic expression

$$dx'^2 + dy'^2 + dz'^2 - c^2dt'^2$$

<sup>1</sup> It is a great convenience, except where expressly stated, to imply that where a suffix occurs twice in an expression a summation is to be effected over the values 1, 2, 3, 4 of that suffix.



has a special physical significance, in relation to the propagation of light and the motion of a free particle. In this case all that has been said in the earlier part of this book would hold. But if we are now dealing with the case of a field that cannot be so transformed away, all that is left of our hypotheses is that there may be an invariant quadratic form

$$g_{rs} dx_r dx_s$$

(we introduce here the letter  $g_{rs}$  instead of  $a_{rs}$  in conformity with the customary notation)—in which the coefficients  $g_{rs}$  are functions of the co-ordinates, and depend upon the particular set of co-ordinates used. This expression has a value independent of the particular set of co-ordinates, and to that extent is capable of representing some metrical property of the physical universe which is independent of any particular mode of observation and method of measurement. But *a priori* nothing is to be assumed as to any limitation upon the coefficients  $g_{rs}$ . These coefficients are characteristic of the particular mode of measurement adopted. The important thing to notice here is that inasmuch as  $g_{rs}$  characterizes the system of reference from which the whole universe is regarded, *there will be no class of phenomena which will not be dependent upon this tensor*. In other words, we have here a basis from which we shall be able to consider the influence of the gravitational field, not only upon the motions of particles, but upon the propagation of light, the motions of electrons, or of any other property of matter we may select.

#### 77. EINSTEIN'S PROBLEM.

We have so far proceeded mainly by way of giving up definite results which classical theory had regarded as fundamental.

It would seem that such a thoroughgoing jettison of all definite hypotheses would leave little hope of arriving at anything in the way of a precise physical law. Let us note therefore, in a very much simpler instance, how what appears to be a negative hypothesis actually suggests a known law. Returning to Newtonian conceptions, let us think of a scalar gravitational potential  $\phi$  with a definite value at each point of space, and ask ourselves what possible connections may

exist between such a potential and the density of matter  $\rho$  also defined for each point of space—on the single hypothesis that space is isotropic, that is, that all directions at a point are alike; or, in other words, on the hypothesis that the form of all physical laws remains the same no matter how the co-ordinate system is changed provided only that  $dx^2 + dy^2 + dz^2$  is invariant.

The problem is thus: we are given two quantities  $\rho$  and  $\phi$  at every point; what types of relation between these two quantities could possibly hold without reference to any special directions in space?

If we had been thoroughly brought up in the analysis of vectors we should at once remember that there is a scalar quantity  $\nabla^2\phi$  which has no relation to any special direction at a point; and at once, if we begin to look for the simplest relations that may hold between  $\phi$  and  $\rho$ ,

$$\nabla^2\phi = k\rho$$

suggests itself as one possibility.

This then suggests the method of approach adopted by Einstein towards this gravitation problem.

Here is the universe, mapped out according to the arbitrary choice of an observer as a space-time field. This observer perceives or interprets his physical universe as having certain gravitational properties characterized by the set of quantities  $g_{rs}$ , and as occupied by a certain distribution of moving matter, with energy and momentum.

His problem is to envisage the possible relations between the quantities  $g_{rs}$  and the matter distribution.

He has to guide him:—

- (i) The gravitational theory of Newton, extremely satisfactory as far as it goes, in which the gravitational field is represented by a single potential satisfying

$$\nabla^2\phi = -4\pi\rho.$$

This at once suggests that the relations sought must be, at the simplest, differential equations of the second order.

- (ii) The criterion that whatever relations are chosen must be equally valid whatever may be the arbitrary manner of depicting the universe in terms of space and time.

With regard to the preservation of the form of the relations in an arbitrary distortion of the four dimensional map, we note that the set of quantities  $g_{rs}'$  will in such an arbitrary distortion differ from the corresponding quantities say  $g_{rs}$ , at the corresponding point. Likewise their differential coefficients with respect to the co-ordinates will take new values.

The problem then is to examine what types of relation connecting the quantities  $g_{rs}$  and their differential coefficients, will represent exactly the same quantitative results as the same relation connecting the quantities  $g_{rs}'$  and their differential coefficients with respect to the new co-ordinates.

For the solution, the necessary pure mathematics was fortunately to hand. We cannot do more here than refer to its originators.<sup>1</sup> Christoffel (1869) attacked this specific problem, and his work forms the foundation of the necessary calculus. Riemann, however, with his peculiar genius, gave significance to the work of Christoffel by showing its intimate relation to the geometry of surfaces, and by his investigation of the possibilities of a geometry in more than three dimensions. This will form the subject of the next section, but we must acknowledge here the debt due to Ricci and Levi-Civita who systematized the work of these two pioneers and left it in a form ready for the physicist to use.

#### 78. GEOMETRICAL—RIEMANN'S CONTRIBUTION.

In 1854 Riemann submitted to the University of Gottingen a thesis "On the Hypotheses which lie at the Bases of Geometry," in which he arrived for questions of measurement in three dimensions at very much the same position as that at which we have now arrived in problems of motion. In order

<sup>1</sup> For an introduction to the mathematical literature of the whole subject, see J. E. Wright, "Invariants of Quadratic Differential Forms," Cambridge Tracts in Mathematics, No. 9.

E. B. Christoffel, "Crelle's Journal für die Math.," LXX., 1869.

B. Riemann, "Über die Hypothesen welche der Geometrie zugrunde liegen," Math. Werke (1876), p. 272.

"Commentatio Mathematica," Math. Werke, p. 370.

The former paper was translated by W. K. Clifford and published in "Nature," 1873.

Ricci and Levi-Civita, "Methodes de calcul différentiel absolu," Mathematische Annalen, 54, 1901.

to extract the logical essence of geometrical relations, stripped of all preconceptions which may possibly be based on muscular sensations of extent, ocular sensations of symmetry and the like, he seeks to reduce geometry first to an arithmetical basis. This is exactly what mathematical physics seeks to do. A point in three-dimension space is for him merely a set of three numbers. The distance between two points is merely some expression in terms of the differences of corresponding numbers which may act as a measure of the separateness of those points. The only property that he demands of this measure is that it should vanish when the two points coincide, and should be essentially a positive quantity if the two points do not coincide (cf. what has been said above as to coincidences in physical measurements). The simplest type of expression satisfying this criterion for two near points he takes to be  $ds$ , the positive square root of  $a_{rs}dx_rdx_s$ , the coefficients  $a_{rs}$  being limited only by the condition that the expression is always positive.

The problem now arises, however, as to what further limitations upon these coefficients may be necessary as a consequence of the fact that all the points considered connect together to form a continuous whole.

As an illustration let us think for a moment of four points of some two-dimensional surface and their six mutual distances. If the surface is plane any five of these distances determine the relative positions of the four points and hence determine the sixth distance. Thus, if the six distances are arbitrarily assigned, the four points will not in general be coplanar.

Extending this, if a surface is defined, and the position of any point upon it is marked by two variable quantities  $\lambda$ ,  $\mu$ , the distance between two consecutive points being given by

$$ds^2 = A d\lambda^2 + 2B d\lambda d\mu + C d\mu^2.$$

A, B, C being arbitrary functions of  $\lambda$  and  $\mu$ , it will not in general be possible that the points should piece together into a plane surface; though there will in general be an infinite number of curved surfaces into which they will piece together. It is possible to write down an expression involving the quantities A, B, C, and their differential coefficients with regard to  $\lambda$  and  $\mu$ , the vanishing of which for all values of  $\lambda$

and  $\mu$  is a necessary and sufficient condition that a plane surface may be so constructed. If this condition is not satisfied, this quantity necessarily has the same value at corresponding points of all possible surfaces. It is called the "specific curvature".

In exactly the same way if we are dealing with a three-dimensional region, the six distances of four points in ordinary space are arbitrary and unrelated, but if a fifth point be added only three of its distances from the other four can be chosen arbitrarily, that is, a certain identical relation must hold between the ten mutual distances. As a consequence of this it is found that an expression  $ds^2 = a_{rs} dx_r dx_s$  cannot, as a rule, satisfactorily represent the distance between two points if they are to piece together into a continuous three-dimensional construct (a so-called Euclidean space).

Riemann in fact is able to write down six independent functions of  $a_{rs}$  and their first and second differential coefficients with regard to the co-ordinates which must vanish if this is to be so. For convenience we will give them his symbolism:  $B_{\mu\nu\sigma}^{\rho} (\mu, \nu, \sigma, \rho = 1, 2, 3)$ . Of the equations  $B_{\mu\nu\sigma}^{\rho} = 0$  only six are independent.

It is possible, however, from Riemann's arithmetical point of view, to think of them piecing together into a three-dimensional region which is, so to speak, a curved region in a four-dimensional world.

In such a region a system of geometry can be built up using the given expression for the element of distance, just as we may have a two-dimensional geometry on the surface of a sphere. Some of the metrical results of Euclidean geometry will cease to be true. For example, the three angles of a triangle on a sphere are not together equal to two right angles. It remains true on a sphere that the angles at the base of an isosceles triangle are equal, and that two triangles whose sides are equal, each to each, are congruent, that is, may be brought into complete coincidence. But neither of these two propositions is true if the triangles are on a surface of which the curvature is not the same at all points, as, for instance, an egg-shaped surface.

A being whose knowledge was limited to the perception of objects in such a two-dimensional universe, discovering this fact of lack of congruence and having no conception of three dimensions, could only say either that all objects were limited in their freedom of movement, or else that their dimensions were altered when they were displaced from one position to another according to some local properties of the places in which they were put. Such alteration might reasonably be interpreted by such a being as due to some field of force. We who live in three dimensions on the other hand ascribe it simply to the limitations of the observer's universe.

#### 79. EXTENSION TO A HIGHER NUMBER OF DIMENSIONS.

Our ordinary systems of pure geometry based on a general notion of absolute direction and absolute length express the properties of surfaces in a form which is independent of any particular choice of numerical co-ordinates. In reducing his geometry to a purely arithmetical basis, Riemann has therefore to replace the intuitive geometrical notions by a mathematical proof that the equations developed represent the same truths whatever be the system of co-ordinates adopted, so that no one system is to be preferred to another. This is completely parallel with the Principle of Relativity, which extends this to all the phenomena of motion.

Having reduced geometry to a numerical basis the way is clear for an extension of the number of dimensions, and Riemann proceeds to the discussion of this. Any set of co-ordinates being  $x_1, x_2, x_3, x_4$ , and the element of length being given by

$$ds^2 = a_{rs} dx_r dx_s,$$

the condition that the elements may fit together into a four-dimension space is again expressed by the equation  $B_{\mu\nu\sigma}^{\rho} = 0$ , but there are now twenty independent expressions instead of six. If these expressions ( $B_{\mu\nu\sigma}^{\rho}$ ) do not vanish two interpretations are open to the observer. The first is that there is a fifth dimension, and that for some reason his four-dimensional universe is a curved section of this, the quantities  $B_{\mu\nu\sigma}^{\rho}$  being measures of the curvature. Or on the other hand if he has no

power to think into five dimensions, he will naturally attribute properties varying from point to point of his four-dimensional world which will correspond to the existence of physical phenomena in that world. As Riemann puts it, "we must seek the ground of its measure relations, in binding forces which act upon it".

If they do vanish Riemann shows that there is a system of co-ordinates in which the element of length is the ordinary expression

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$$

or, in geometrical language the measure relations are those of Euclidean space,  $x_1, x_2, x_3, x_4$ , being then a set of rectangular co-ordinates.

#### 80. APPLICATION TO PHYSICAL THEORY, CONDITIONS FOR ABSENCE OF REAL GRAVITATIONAL FIELD.

Without using any of the geometrical language of the preceding section, which has been inserted by way of illustration, we may quote this main result of the mathematical theory.

*There are twenty independent combinations of the quantities  $g_{rs}$  and their first and second differential co-efficients with respect to the co-ordinates, the vanishing of which gives the conditions that the element  $ds^2$  can by a suitable change of co-ordinates be put in the form*

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2.$$

If this is the case, and an observer used these co-ordinates, a free particle would appear to describe a straight line uniformly and light would travel in straight lines—there would in fact be no evidence of any gravitational effect.

These quantities being denoted by  $B_{\mu\nu\sigma}^p$  the equations

$$B_{\mu\nu\sigma}^p = 0$$

*express then the condition for the absence of a real gravitational field.*

#### 81. EXTENSION TO INCLUDE A REAL GRAVITATIONAL FIELD.

Now our ordinary experience is that gravitation and matter are intimately allied with one another, and we shall not expect

that we shall be able in the material universe to choose a frame of co-ordinates which shall entirely transform away the gravitational field. In the case therefore where the functions  $B_{\mu\nu\sigma}^\rho$  do not all vanish, the question arises as to how they are to be related to the presence of matter in the field.

Looking back to the familiar Laplace equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0$$

and having come to the conclusion that the gravitational field must be described in terms of the tensor  $g_{\mu\nu}$  instead of the single quantity  $\phi$ , Einstein looks for a set of equations in number equal to the number of  $g$ 's, differential equations of the second order, which are satisfied in the particular case of a purely fictitious field, that is when  $B_{\mu\nu\sigma}^\rho = 0$ , but which are more general in character. These equations must of course retain their form under any arbitrary change of co-ordinates.

Putting  $G_{\mu\nu} = \sum_\rho B_{\mu\nu\rho}^\rho$  ( $\rho = 1, 2, 3, 4$ ), it is found that *the set of equations*

$$G_{\mu\nu} = 0$$

*satisfies all the necessary criteria for being a suitable extension of the equation  $\nabla^2 \phi = 0$ , the equation satisfied by the gravitational field at a point not actually occupied by matter.*

It is shown that if a change of co-ordinates be made whereby  $g_{\mu\nu}$  change into  $g'_{\mu\nu}$ , and  $G_{\mu\nu}$  into  $G'_{\mu\nu}$  then the equations connecting  $G'_{\mu\nu}$  with  $G_{\mu\nu}$  are identical with those connecting  $g'_{\mu\nu}$  with  $g_{\mu\nu}$ . Such a set of quantities is said to be a *covariant tensor* of the second order. Einstein seizes upon these equations, and takes them as characteristic of a region not occupied by matter. We shall see later how he was able to make use of them.

### *Summary of Mathematical Formulation.*

$$ds^2 = g_{rs} dx_r dx_s \quad . \quad . \quad . \quad (i)$$

If

$$b_s = g_{rs} a^r, \quad s = 1, 2, 3, 4$$

and these equations are solved to give  $a^s$  in terms of  $b_r$ , we write

$$a^s = g^{rs} b_r \quad . \quad . \quad . \quad (ii)$$

$$[\mu\nu, \lambda] = \frac{1}{2} \left( \frac{\partial g_{\mu\lambda}}{\partial x^\nu} + \frac{\partial g_{\nu\lambda}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\lambda} \right) \quad . \quad . \quad (iii)$$

$$\{\mu\nu, \lambda\} = g^{\lambda\alpha} [\mu\nu, \alpha] \quad . \quad . \quad . \quad (iv)$$



$$B_{\mu\sigma}^p = \{\mu\sigma, \epsilon\} \{\epsilon\nu, \rho\} - \{\mu\nu, \epsilon\} \{\epsilon\sigma, \rho\} + \frac{\partial}{\partial x^\nu} \{\mu\sigma, \rho\} - \frac{\partial}{\partial x^\sigma} \{\mu\nu, \rho\} \quad (v)$$

$$G_{\mu\nu} = B_{\mu\nu\rho}^p \quad (vi)$$

As always, it is implied that summations are to be affected with regard to the suffixes which occur twice in all these expressions, viz. :—

$r$  in (ii),  $a$  in (iv),  $\epsilon$  in (v), and  $\rho$  in (vi).

## 82. THE PATH OF A FREE PARTICLE HAS STATIONARY LENGTH.

Although Einstein's guess at a possible form of the equations was actually arrived at along the lines just indicated, it will perhaps enable the reader to get a firmer hold of the fact that a set of differential equations expressed in terms of a special set of co-ordinates in reality embodies a fact entirely independent of the interpretation in terms of space and time which is attached to those co-ordinates, if we express them in a different but equivalent form.

In Newtonian dynamics we know that a particle in a field in which there is no force acting on it moves in a straight line uniformly. In the four-dimensional picture of its motion in which the Newtonian space and time co-ordinates are actually used, the motion of the particle is represented by a straight line, that is, a line such that the differences in the space co-ordinates of two positions of the particle are proportional to the difference in the time co-ordinate.

We recall the identification of a straight line as the shortest distance between two points. The length of a line in a four-dimensional picture is to be defined thus. In the absence of a gravitational field the interval between two adjacent points has been defined in terms of the small differences of co-ordinates, thus—

$$ds^2 = -dx_1^2 - dx_2^2 - dx_3^2 + dx_4^2.$$

Let a series of points be taken along the curve and the sum of the intervals between each point and the next be calculated. We find the limit to which this sum tends as the subdivision of the line becomes finer and finer, and we call this sum the total interval between the terminal points for the given curve.

Denoting this by  $\int ds$ , it is clear that if the curve is changed, the length of it is in general changed. But let the curve be straight, in the sense spoken of above, and then let it be bent a little out of the straight by giving to each point of it an arbitrary small displacement. The mathematician can show, by what is known as the calculus of variations, that such a slight displacement will, to the first order of small quantities, leave the total length of the curve unaltered. We say that it has a stationary value. Further, it is capable of proof that the straight line is the only curve for which this is true. This is the equivalent of the shortest length property of the Euclidean straight line.

Thus the path of a free particle in the absence of any gravitation is such that  $\int ds$  is stationary, for any slight variation from its passage between two given point instants. For convenience the term 'geodesic' may be used to connote this property, a term borrowed from the geometry of curves on surfaces. But we do not conclude that the total interval is a true minimum. As a matter of fact it is in general a maximum.<sup>1</sup>

### 83. THE GEODESIC PROPERTY A RELATIVE ONE.

Consider the picture of the same particle in the same conditions merely by transforming co-ordinates. In the new picture there is what we have spoken of as a *fictitious* gravitational field. The particle will freely obey this field, but its world-line will no longer be straight. But, inasmuch as there is an exact point-to-point correspondence between the two pictures, and the separation of two adjacent points  $ds$  is the same in both, the path in the transformed picture will be also a geodesic.

At this point we recall what has been said of the equivalence of a fictitious and essential gravitational field. Thus we pass naturally to the hypothesis that *the path of a particle in any field is such that the total separation of the initial and final configurations is stationary.*

<sup>1</sup> Because we have taken  $ds^2 = -dx_1^2 - dx_2^2 - dx_3^2 + dx_4^2$   
instead of  $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$ .

This brings us to the mathematical problem of finding the path for which

$$\int (g_{rs} dx_r dx_s)^{\frac{1}{2}}$$

is stationary as the equivalent of the problem of finding how a free particle will move in a field of gravitational force characterized by the tensor  $g_{rs}$ .

So far, of course, the precise relation of this tensor to the ordinary gravitational potential has not been shown. This will appear later when we take some particular cases.

In the same way, in the complete absence of a gravitational field, real or fictitious, we come back to the equation of propagation of a ray of light as

$$ds = 0.$$

In a purely fictitious field the same equation must be satisfied, and, in a field which is not entirely fictitious, the assumption that the same equation holds will satisfy the hypotheses of Einstein. It remains to be seen whether it agrees with experiment.

#### 84. HAMILTON'S PRINCIPLE OF STATIONARY ACTION.

In passing to the wider field of physical phenomena in general, we take up the suggestions from various quarters<sup>1</sup> that, in the re-thinking of physical theory, the Principle of Least Action is that which stands least in need of modification.

We have seen in regard to the special Principle of Relativity that that Principle not only left the action principle standing—but that the action was an invariant quantity, in fact the only quantity associated with material systems which had a value independent of the space-time system of measurement.

We naturally, therefore, look for a quantity which is invariant under the most general transformation, and, having found one, examine whether the consequences of treating this quantity as the action will lead to laws at all consonant with the known phenomena.

The researches of Riemann give us at once a quantity that is invariant, namely, the quantity  $G = \sum_{\mu} \sum_{\nu} g^{\mu\nu} G_{\mu\nu}$ , the summation being made over all possible sets of values of  $\mu, \nu$ . This is a quantity defined for each point of the four-dimensional world (that is, of time and space).

<sup>1</sup> Particularly by Sir Joseph Larmor, "Æther and Matter," and elsewhere.

85. NOTE ON THE GEOMETRICAL SIGNIFICANCE OF  $G$ .

Here the work of the geometers, referred to above (§ 78), is of fundamental importance. Gauss showed that if a surface of inextensible but flexible material be deformed—the geometrical conditions being merely that the distance between two neighbouring points does not alter, then throughout all the changes of shape that the surface may undergo there is one quantity associated with each point of the surface that does not change. He called it the *specific curvature*; in the language of the geometry of surfaces it is the product of the curvatures of the principal sections of the surface at the point. Its value is determined simply by the nature of the surface, and is, therefore, quite unconcerned with any particular mathematical mode of defining the position of a point on the surface.

This can be generalized, following up the lead given by Riemann to any number of dimensions. For our purpose here it is sufficient to know that *in any four-dimensional connected region built up of fixed measured elements of fixed length there is an invariant quantity associated with each point*—we may call it the *generalized curvature*—though that phrase has no meaning except in relation to a hypothetical five-dimensional space in which the four-dimensional region is drawn as an ordinary surface is drawn in three-dimensional space. This quantity depends only on the mutual relations between the lengths of neighbouring elements, and not upon any particular analytical method of specification. It is thus that the invariant quantity  $G$  is reached.

## 86. INVARIANT ELEMENT OF INTEGRATION.

In § 62 in which the electro-magnetic field was being considered in relation to the principle of least action, it was remarked that under the Lorentz-Einstein transformations, the element of integration in the four-dimensional world was invariant. This was clear because those transformations consisted merely in regarding the same element from a set of co-ordinates obtained from the old by a mere turning round of the axes.

In the present general theory, in order to compare the

contents of the given element as portrayed in two separate sets of co-ordinates, we have to remember that we not only turn the axes about but we may stretch and warp the picture. We find, however, that the element of the four-dimensional space is increased in exactly the same ratio that  $\sqrt{g}$  is diminished;<sup>1</sup> so that there is an invariant element  $\sqrt{g} dV$ .

### 87. HYPOTHETICAL EXPRESSION FOR ACTION.

We thus arrive at an expression which for any given region of the four dimensional world is an invariant, viz:—

$$\int G \sqrt{g} dV.$$

The subject of integration here is simply a function of the quantities  $g_{rs}$  and their first and second differential coefficients in regard to the co-ordinates.

For a field in which gravitation can be completely transformed away, a field which has previously been called a Galilean field, the gravitation being as we have said completely fictitious,<sup>2</sup> the value of  $G$  is zero.

The conditions that this integral shall not change its value (to the first order) when arbitrary small changes are given to the  $g$ 's, are found to be exactly that set of differential equations upon which Einstein first alighted as the simplest with which the repertory of the pure mathematicians could furnish him, viz:—

$$G_{\mu\nu} = 0.$$

Thus we are able to put Einstein's theory in this comprehensive form, that *there is a numerical quantity which can be calculated out of the distribution of the gravitational field, the value of which is entirely independent of the choice of the observer as to the means by which he describes the positions and motions of the material bodies; and that the distribution of the field is always such as to make this value stationary as compared with the value of any slightly differing field.*

<sup>1</sup>  $g$  stands for the determinant

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix}$$

## 88. THE MATTER TENSOR.

Returning to the explicit differential equations adopted by Einstein, it is necessary to contemplate what happens at points at which  $G_{\mu\nu}$  does not vanish. If  $G_{\mu\nu} = 0$  is the equation characteristic of the absence of matter, the expression  $G_{\mu\nu}$  must measure the relevant properties of matter. Our conception of those properties may have to be revised to meet the new theory. We have already learned the interdependence of energy and momentum, and the relative nature of force and stress. We can therefore only say at present that there must be a relation of the form

$$G_{\mu\nu} = \Theta_{\mu\nu}$$

where  $\Theta_{\mu\nu}$  is some tensor which is characteristic of the state of matter at a point. The fact that it is a tensor will ensure that the equation represents a physical relation independent of the choice of the system of co-ordinates.

We speak of  $\Theta_{\mu\nu}$  as the *matter tensor* or the *energy tensor*. Its relation to the motion of the matter and the ordinary conceptions of the properties of matter may for the moment be left; but it should be emphasized that the above equation does no more than identify certain aspects of the gravitational field with the matter. It associates together the curvature of the measure system as defined by the quantities  $G_{\mu\nu}$  and the life history of material bodies.

## 89. THE LAWS OF CONSERVATION IN EINSTEIN'S THEORY.

In the old mechanics, three laws stand out prominently, the laws of conservation of mass, of momentum, and of energy. In Einstein's early theory of relativity, and indeed prior to that, the conservation of mass ceased to have validity; we have seen that the idea of constant mass is intimately associated with that of constant internal energy (p. 80).

But the laws of conservation of momentum and energy remain in Einstein's theory; in fact they can be shown to be an immediate consequence of his law of gravitation. The relation of these laws of conservation to the general laws of

the field may be put thus. Starting from the expression for the action,

$$A = \int G \sqrt{g} d\mathcal{V},$$

an integral through all space of a function of  $g_{\mu\nu}$  and its differential coefficients, two conditions have to be satisfied.

- (i)  $A$  is invariant when a change of co-ordinates is made with a corresponding change in  $g_{\mu\nu}$
- (ii)  $\delta A = 0$  when arbitrary changes  $\delta g_{\mu\nu}$  are given to the tensor  $g_{\mu\nu}$ .

Now it is quite easy to see that the conditions that (ii) may be so, include the conditions for (i). The condition for (ii) is in fact Einstein's equation

$$G_{\mu\nu} = 0.$$

If  $G$  were not known to be invariant these would be ten independent equations, but if  $G$  is invariant there are four expressions derivable from the ten quantities  $G_{\mu\nu}$  which vanish identically whether  $G_{\mu\nu} = 0$  or not.

Thus when  $G_{\mu\nu}$  is interpreted as characterizing the state of matter at any point, there is found to be a quartet of relations satisfied by these quantities. The form of these relations corresponds precisely to the form of the equations usually taken as representing the laws of conservation of energy and of momentum.

Thus on Einstein's theory these laws are an immediate consequence of the invariance of the action, quite apart from the principle of least action.<sup>1</sup>

<sup>1</sup> For a different expression of the intimate relation between conservation and relativity, see Eddington, "Space, Time, and Gravitation," p. 139.

## CHAPTER VIII.

### THE VERIFICATION OF EINSTEIN'S EQUATIONS.

#### 90. CONSEQUENCES OF THE NEW THEORY.

WE come now to an examination of the two results of Einstein's theory which have been so strikingly verified.

Newton's simple law of the inverse square for gravitation together with his scheme of dynamics had proved to account with great accuracy for all the planetary and lunar motions with all the irregularities observed in those motions,<sup>1</sup> except for one or two outstanding instances, of which the most marked was that of the motion of the perihelion of Mercury. After allowance for all known causes there remained an unexplained progression at the rate of  $40\cdot2''$  per century.

In 1915 Einstein was able to show that his new theory would in the first place give all the results of the Newtonian theory; but further he found that a correction had to be added which in the case in question amounts to a progression at the rate of  $43''$  per century. In all other cases the correction is negligible and leaves the agreement between theory and observation unspoiled.

It seems sufficiently surprising that generalities such as those which have been described in the preceding pages should suggest any precise numerical result at all; but that they should succeed in explaining the one outstanding discrepancy within the degree of accuracy of which the observations are capable is sufficient to bring the theory at once into the front rank.

The second verification of Einstein's theory was in respect of its prediction that light rays from a star would be deflected if they pass near to the sun, thus producing a displacement

<sup>1</sup> The classical instance is the prediction of the existence of the planet Neptune from a consideration of the perturbations of Uranus by Prof. J. Couch Adams and Leverrier.



which should be measurable. This we shall consider in § 93. But first we must consider in detail Einstein's discussion of the field of gravitation round the sun.

#### 91. THE FIELD ROUND THE SUN TREATED AS A SINGLE GRAVITATING PARTICLE.

The law which Einstein proposed for the gravitational field in free space is, as has been said,

$$G_{\mu\nu} = 0.$$

This is a set of ten equations connecting the quantities  $g_{\mu\nu}$  and their first and second differential coefficients.

We begin to ask what results from this differential equation if the field is associated with a single centre, thinking of the sun as a mere particle or as a spherical distribution of matter.

Two characteristics may at once be laid down :—

- (i) The field is constant, that is does not vary with the time.
- (ii) The sun's gravitational field as measured by an observer on the sun, will be symmetrical round its centre.

The first condition requires that the quantities  $g_{\mu\nu}$  are all independent of the time.

The second means that the expression

$$g_{rs} dx_r dx_s$$

must be built up from quantities which have no special relation to any particular direction through the centre of the sun.

If we use spherical polar co-ordinates the only such quantity in addition to  $r$ , the distance from the centre and quantities derived from it, is the element of distance between two neighbouring points on a sphere whose centre is at the centre of the sun. The square of this distance is known to be

$$r^2(d\theta^2 + \sin^2\theta d\phi^2).$$

Thus the two conditions limit the form of  $ds$  to that given by

$$ds^2 = f(r) dr^2 + \phi(r) (d\theta^2 + \sin^2\theta d\phi^2) + \psi(r) dt^2.$$

But again, in view of the general Principle of Relativity, there is nothing to say what is the proper way to measure the distance from the centre of the sun ; so that we will say quite arbitrarily that  $r$  shall be so measured that  $\phi(r)$  is  $-r^2$ .

Thus  $ds^2$  is now

$$f(r) dr^2 - r^2(\sin^2\theta d\phi^2 + d\theta^2) + \psi(r) dt^2.$$

Or,  $g_{11} = f(r)$ ,  $g_{22} = r^2 \sin^2 \theta$ ,  $g_{33} = r^2$ ,  $g_{44} = \psi(r)$ ,  
and all the other  $g_{\mu\nu}$  are zero.

It remains to write down the explicit form of the equations

$$G_{\mu\nu} = 0$$

with these simplifications; when this is done it is found that it is possible to find without ambiguity the form of the functions  $f(r)$  and  $\psi(r)$ ; in fact we find that

$$\psi(r) = 1 - A/r \text{ and } f(r) = - (1 - A/r)^{-1}$$

where  $A$  is a constant of integration.  $A$  is in fact the constant which shows how the field differs from a field of no gravitation; for if  $A$  were zero we should have

$$ds^2 = - dr^2 - r^2 (\sin^2 \theta d\phi^2 + d\theta^2) + dt^2,$$

which in ordinary Cartesian co-ordinates would become

$$- dx^2 - dy^2 - dz^2 + dt^2.$$

Thus, except for the evaluation of the constant  $A$ , there is a unique solution of the gravitational equations subject to the conditions of permanence and symmetry. This solution applies only to the region from which matter is absent. It would hold in the region outside a spherical distribution of matter, but not within the matter itself.

The constant  $A$  clearly has something to do with the sun's power to produce a gravitational field; and we shall see that it corresponds to  $2\gamma m$ , where  $m$  is the mass of the sun, and  $\gamma$  the constant of gravitation.

## 92. THE MOTION OF A PLANET IN THIS FIELD.

The next step is the determination of the motion of a particle moving freely in this field; a motion which is determined by the condition that  $\int ds$  has a stationary value. The world-line of the particle is a shortest distance in the four-dimensional map.

We take any arbitrary world-line from a point  $O$  to a point  $Q$ .

Let  $s = \int_0^P ds$ ,  $P$  being any

point on this line. Let an adjacent world-line be drawn, the point  $P'$  corresponding to  $P$  being such that its co-ordinates

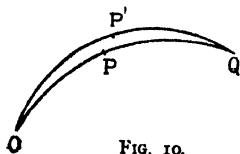


FIG. 10.

differ from those of  $P$  by  $\delta r$ ,  $\delta \theta$ ,  $\delta \phi$ ,  $\delta t$ , these being arbitrary functions of  $s$ , subject to the condition that they vanish at the ends.

Then the increment in  $\int_0^s ds$  when evaluated along this varied line instead of the original line is

$$\delta \int \left\{ \psi dt^2 - dr^2/\psi - r^2(\sin^2 \theta d\phi^2 + d\theta^2) \right\}^{\frac{1}{2}} \\ = \delta \int F ds$$

where  $F^2 = \psi dt^2 - r^2/\psi - r^2(\sin^2 \theta \phi'^2 + \theta'^2)$ , accents denoting differentiations in regard to  $s$ . According to the ordinary method of the calculus of variations<sup>1</sup> this change will vanish to the first order in  $\delta t$ ,  $\delta r$ ,  $\delta \theta$ ,  $\delta \phi$ , whatever functions of  $s$  these may be if

$$\begin{aligned} \text{(i)} \quad \frac{d}{ds} \left( \frac{\partial F}{\partial r'} \right) &= \frac{\partial F}{\partial r} & \text{(ii)} \quad \frac{d}{ds} \left( \frac{\partial F}{\partial \theta'} \right) &= \frac{\partial F}{\partial \theta} \\ \text{(iii)} \quad \frac{d}{ds} \left( \frac{\partial F}{\partial \phi'} \right) &= 0 & \text{(iv)} \quad \frac{d}{ds} \left( \frac{\partial F}{\partial t'} \right) &= 0. \end{aligned}$$

From (iii) and (iv)  $\frac{\partial F}{\partial \phi'}$  and  $\frac{\partial F}{\partial t'}$  are constant along the required path. Also by the definition of  $ds$ ,  $F = 1$  all along this path.

Thus we obtain

$$r^2 \sin^2 \theta \phi' = h$$

and

$$\psi t' = k$$

$h$  and  $k$  being constants.

<sup>1</sup> Thus

$$\delta \int F ds = \int \delta F ds = \int \left\{ \frac{\partial F}{\partial r} \delta r + \frac{\partial F}{\partial \theta} \delta \theta + \frac{\partial F}{\partial r'} \delta r' + \frac{\partial F}{\partial \phi'} \delta \phi' + \frac{\partial F}{\partial \theta'} \delta \theta' + \frac{\partial F}{\partial t'} \delta t' \right\} ds.$$

Also

$$\delta r' = \delta \frac{dr}{ds} = \frac{d}{ds} \delta r$$

so that, integrating by parts,

$$\int \frac{\partial F}{\partial r'} \delta r' ds = \int \frac{\partial F}{\partial r'} \frac{d}{ds} (\delta r) ds = \left[ \delta r \frac{\partial F}{\partial r'} \right]_0^s - \int \frac{d}{ds} \left( \frac{\partial F}{\partial r'} \right) \delta r ds = - \int \frac{d}{ds} \left( \frac{\partial F}{\partial r'} \right) \delta r ds,$$

if  $\delta r = 0$  at both ends of the path. Similarly for the other co-ordinates. Thus

$$\delta \int F ds = \int \left\{ \delta r \left[ \frac{\partial F}{\partial r} - \frac{d}{ds} \left( \frac{\partial F}{\partial r'} \right) \right] + \delta \theta \left[ \frac{\partial F}{\partial \theta} - \frac{d}{ds} \left( \frac{\partial F}{\partial \theta'} \right) \right] \right. \\ \left. - \delta \phi \frac{d}{ds} \frac{\partial F}{\partial \phi'} - \delta t \frac{d}{ds} \frac{\partial F}{\partial t'} \right\} ds.$$

Again 
$$F \frac{\partial F}{\partial \theta'} = -r^2 \theta'$$

$$\therefore F \frac{d}{ds} \left( \frac{\partial F}{\partial \theta'} \right) + \frac{dF}{ds} \frac{\partial F}{\partial \theta'} = - \frac{d}{ds} (r^2 \theta')$$

or, since  $F = 1$  and therefore  $\frac{dF}{ds} = 0$

$$\frac{d}{ds} \left( \frac{\partial F}{\partial \theta'} \right) = \frac{d}{ds} (r^2 \theta')$$

Also 
$$F \frac{\partial F}{\partial \theta} = -r^2 \sin \theta \cos \theta \phi'^2$$

$$\therefore r^2 \sin \theta \cos \theta \phi'^2 = r^2 \theta'' + 2r r' \theta'$$

Now, at any specified moment the particle is moving in a certain plane through the origin. Let the axis from which  $\theta$  is measured be taken normal to this plane. Then at this moment  $\theta = \frac{\pi}{2}$  or  $\cos \theta = 0$ , and  $\theta' = 0$ .

Hence  $\theta''$  is zero, and  $\theta$  remains constantly at the value  $\pi/2$ . Thus the orbit remains continually in this plane.

Instead of integrating equation (i) we use  $F = 1$ , which must be consistent with it.

Putting  $\theta = \pi/2$  and  $\psi l' = k$  this becomes

$$r'^2 / \psi + r^2 \phi'^2 = -1 + k^2 / \psi.$$

Substituting  $r^2 \phi' = h$

this gives  $r'^2 = k^2 - 1 + A/r - h^2/r^2 + Ah^2/r^3$ .

In the absence of the last term this equation would have been exactly the Newtonian equation, except that here  $r'$  is  $dr/ds$  instead of  $dr/dt$ . As far as the shape of the orbit goes, however, this makes no difference, since it is obtained by eliminating  $ds$  or  $dt$  between the last two equations to give an equation connecting  $r$  and  $\phi$  only.

Writing  $A = 2\gamma m$ ,  $h^2 = \gamma m l$ , and  $k^2 - 1 = \gamma m/a$  the last equation becomes

$$r'^2 = \gamma m \left( -\frac{1}{a} + \frac{2}{r} - \frac{l}{r^2} + \frac{2\gamma m l}{r^3} \right).$$

The corresponding Newtonian equation is

$$(dr/dt)^2 = \gamma S \left( -\frac{1}{a} + \frac{2}{r} - \frac{l}{r^2} \right)$$

where  $S$  is the mass of the sun.

Thus we are able to identify  $m$  with the mass of the sun.

## 93. NUMERICAL CALCULATION OF THE CORRECTION TO NEWTONIAN THEORY.

The value of  $\gamma m$  (taking the unit of length to be 1 kilometre, and the unit of time  $\frac{1}{300000}$ th second, so that the velocity of light is unity) is given by the equation

$$\text{periodic time } T = 2\pi a^{\frac{3}{2}} / \sqrt{\gamma m}$$

which from the earth's orbit gives

$$\gamma m = a^3 / 4\pi^2 T^2 = 1.47.$$

Thus, in the orbit of Mercury, since  $r$  is of order  $10^8$  (kilometres)  $\gamma m/r$  is of the order  $10^{-8}$ .

Thus  $\psi$  differs from 1 by the order  $10^{-8}$ , and the extra non-Newtonian term (in our equation) is of the same order compared with the other terms on the right-hand side. We may say, therefore, that to that order of approximation the Newtonian calculations are correct.

We proceed to find the effect on the orbit of the correcting term, taking account only of the first order in  $\psi$ .

The equation for the orbit eliminating  $s$  or  $t$  is

$$r^2 \left( \frac{d\phi}{dr} \right)^2 = \frac{h^2}{\gamma m \left( -\frac{1}{a} + \frac{2}{r} - \frac{l}{r^2} + \frac{2\gamma m l}{r^3} \right)}$$

or putting as usual  $u = 1/r$ .

$$\left( \frac{d\phi}{du} \right)^2 = \frac{1}{-1/al + 2ul - u^2 + 2\gamma mu^3}.$$

This is to be integrated to give the polar equation of the orbit.

Here we consider only the first order effect on the apse lines due to the additional term.

Write the equation

$$\left( \frac{d\phi}{du} \right)^2 = \frac{1}{(a-u)(u-\beta)(B-2\gamma mu)}$$

where  $a$  and  $\beta$  are the corrected values of  $u$  for which  $du/d\phi = 0$ , that is, the new apsidal distances.  $B$  is given by

$$B + 2\gamma m(a + \beta) = 1,$$

and, therefore, is only slightly different from 1.

During the motion  $u$  is confined between  $a$  and  $\beta$ .

$$\begin{aligned} \text{Thus } \frac{d\phi}{du} &= \frac{1}{\sqrt{B(a-u)(u-\beta)}} \left( 1 - \frac{2\gamma mu}{B} \right)^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{B(a-u)(u-\beta)}} \left( 1 + \frac{\gamma mu}{B} + \dots \right) \end{aligned}$$

We integrate between  $u = a$  and  $u = \beta$  to obtain the angle  $\phi$  between the apse lines. In the Newtonian case this angle is  $\pi$ .

$$\text{Here } \phi = \frac{1}{\sqrt{B}} \int_a^\beta \frac{du}{\sqrt{(a-u)(u-\beta)}} + \frac{\gamma m}{B^{\frac{1}{2}}} \int_a^\beta \frac{u du}{\sqrt{(a-u)(u-\beta)}} \\ = \frac{\pi}{B^{\frac{1}{2}}} + \frac{\pi \gamma m(a + \beta)}{2B^{\frac{1}{2}}}$$

Putting in the value of  $B$  in the first term, and neglecting  $\{\gamma m(a + \beta)\}^2$

$$\phi = \pi \{1 + \gamma m(a + \beta)\} + \frac{\pi}{2} \gamma m(a + \beta) \\ = \pi \{1 + \frac{3}{2} \gamma m(a + \beta)\}$$

In the second term, which is small, we may put for  $a$  and  $\beta$  their approximate values  $a/(1 - e)$  and  $a/(1 + e)$  of the Newtonian orbit so that

$$(a + \beta) = 2/a(1 - e^2) = 2/l$$

Thus

$$\phi = \pi(1 + 3\gamma m/l).$$

The angular distance from apse to apse exceeds  $\pi$  therefore by the small fraction  $3\gamma m/l$  of itself. Thus the apse line moves forward in the plane of the orbit.

For Mercury we have

$$l = 5.55 \times 10^7 \text{ kilometres, and } \gamma m = 1.47 \text{ kilometres,} \\ \text{giving } 3\gamma m/l = 7.95 \times 10^{-8}.$$

The time of one revolution of Mercury is 87.97 days.

With these figures the rate of advance of the apse line works out as

$$42.9'' \text{ per century.}$$

The outstanding discrepancy to be explained was

$$40.2'' \text{ per century.}$$

#### 94. THE DEFLECTION OF LIGHT IN A GRAVITATIONAL FIELD.

The law which has been suggested for the propagation of light is  $ds = 0$ . Taking the expression obtained above this gives

$$\frac{1}{\psi} \left( \frac{dr}{dt} \right)^2 + r^2 \left\{ \left( \sin^2 \theta \frac{d\phi}{dt} \right)^2 + \left( \frac{d\theta}{dt} \right)^2 \right\} = \psi$$

$$\text{or } \left( \frac{dr}{dt} \right)^2 + r^2 \left( 1 - \frac{2\gamma m}{r} \right) \left\{ \sin^2 \theta \left( \frac{d\phi}{dt} \right)^2 + \left( \frac{d\theta}{dt} \right)^2 \right\} = \left( 1 - \frac{2\gamma m}{r} \right)^2.$$

Confining our attention to a region in which  $\gamma m/r$  is small we can rewrite this equation

$$\left(\frac{dr_1}{dt}\right)^2 + r_1^2 \left\{ \sin^2 \theta \left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\theta}{dt}\right)^2 \right\} = \left(1 - \frac{2\gamma m}{r_1}\right)^2$$

where  $r_1 = r - \mu$ ; so that if we plot the path of the ray of light by means of the co-ordinates  $r_1$ ,  $\theta$ ,  $\phi$ , we find that its velocity is  $1 - \frac{2\gamma m}{r_1}$

In this picture then the light will move as in a medium whose refractive index  $\mu$  is  $\left(1 - \frac{2\gamma m}{r_1}\right)^{-1}$  or  $1 + \frac{2\gamma m}{r_1}$ . This increases as  $r_1$  diminishes so that the rays will be concave to the origin.

To calculate the complete deflection of a ray coming from a long distance and travelling to a great distance, we proceed thus: if  $p$  is the perpendicular from the centre on the tangent the ray travels so that  $\mu p = \text{const.}$

$$\begin{aligned} \text{Thus} \quad p \left(1 + \frac{2\gamma m}{r_1}\right) &= \text{const.} \\ &= R \left(1 + \frac{2\gamma m}{R}\right) \end{aligned}$$

if  $R$  is the distance of nearest approach to the origin.

$$\begin{aligned} \text{This gives} \quad \frac{R(1 + 2\gamma m/R)}{p} &= 1 + \frac{2\gamma m}{r_1} \\ \text{or squaring} \quad \frac{R^2(1 + 4\gamma m/R)}{p^2} &= 1 + \frac{4\gamma m}{r_1}. \end{aligned}$$

This can be identified with the polar equation of a hyperbola with respect to the focus, viz. :—

$$\frac{b^2}{p^2} = 1 + \frac{2a}{r_1}$$

by putting

$$a = 2\gamma m$$

and

$$b^2 = R^2(1 + 4\gamma m/R)$$

or

$$b = R(1 + 2\gamma m/R).$$

The total deflexion of the ray is the angle between the asymptotes, viz. :—

$$2 \tan^{-1} \frac{a}{b}$$

which to the first order is

$$2 \tan^{-1} \frac{2\gamma m}{R} = \frac{4\gamma m}{R}.$$

On drawing the corresponding curve with  $r$  as the radius vector instead of  $r_1$  it is easily seen that it is a curve with the same asymptotes, so that the deflexion in the system of co-ordinates commonly used is that calculated.

Taking  $\gamma m = 1.47$  kilometres and  $R = 697,000$  kilometres we find for the predicted deflexion of a ray passing close to the sun's surface a deflexion of  $1.74''$ .

This figure was strikingly verified by the results of the photographs of stars taken in the region round the sun at the total eclipse in May 29, 1919. An account of the details is to be read in Prof. Eddington's "Space, Time, and Gravitation," Chapter VII. Some further discussion of the interpretation of the result will be found in the next chapter (p. 121).

#### 95. THE SUGGESTED SHIFT OF THE SPECTRUM BY GRAVITATION.

The prominence given to the *interval* between two events as an invariant quantity with a physical significance<sup>1</sup> has led to a much-discussed suggestion that the radiation from a given atom will be altered in period if the atom is moved into a gravitational field of different intensity. The suggestion arises thus.<sup>2</sup> Let  $ds$  be the interval measure between the beginning and end of a single vibration of a radiating atom, as measured in a given set of co-ordinates, at a certain place in a gravitational field. By variation of the system of co-ordinates, we shall see the same atom in an apparently different field of gravitation, but  $ds$  will remain the same. Thus it is suggested that  $ds$  will be a permanent quantity, characteristic of the vibration of this atom, and that all identical atoms will give the same vibration interval  $ds$  whatever the field in which they were placed.

If this were so, if two identical atoms were *at rest* in two different places, their time-space co-ordinates being respectively  $(x, y, z, t)$ ,  $(x', y', z', t')$ , and the gravitational tensor being  $g_{\mu\nu}$  and  $g'_{\mu\nu}$  for the two atoms respectively, we should have

$$g_{44}\delta t^2 = g'_{44}\delta t'^2$$

on putting  $\delta x = \delta x' = 0$ , etc., to express that the atoms are at rest in their respective fields. Thus the corresponding intervals

<sup>1</sup> Cf. § 76.

<sup>2</sup> See Eddington, "Space, Time, and Gravitation," p. 128.



of time, for the two vibrations of the respective atoms, are in the ratio  $\sqrt{g_{44}} : \sqrt{g'_{44}}$ . Taking the values of  $g_{44}$  at the sun's surface, and at the distance of the earth from the sun in the field, as given in § 90, this ratio differs from unity by about '000002, which would correspond to a shift of light of wavelength  $4000 \text{ \AA}$  by an amount  $.008 \text{ \AA}$  towards the red end.

Such evidence as there is, is apparently against the existence of such an effect.

But it is important to note that the argument is based on a hypothesis which has not been used in connection with the two phenomena which have confirmed Einstein's theory; and that is that nature itself will provide a measure of the interval between two events. It is true that in the transformation from one co-ordinate system to another the interval between two given events remains fixed. But that is not the same as saying that, in any one co-ordinate system, we can say of two pairs of events, defined by the vibrations of two widely separated atoms, that the intervals are the same.

Thus the failure to find evidence of this suggested phenomena does not weaken the general confirmation of Einstein's theory. Not only so, but in the next chapter (§ 105) we shall see that a generalization of Einstein's theory is made in which the interval between two events is an indeterminate multiple of the quantity  $(g_{rs}dx_rdx_s)^{\frac{1}{2}}$ , so that it becomes impossible to say that this is a universal quantity for the interval between two successive vibrations of an atom.

## 96. CONCLUSION.

Beginning then with an attempt to satisfy the metaphysician, Einstein has succeeded in making an analysis of physical phenomena which is in most marked agreement with observation. It is therefore an analysis which has to be accepted as the most comprehensive hitherto achieved.

Let us now summarize the position, especially, in regard to the concepts of space and time, emphasizing the point at which the new theory goes further than the special principle of 1905.

Like that principle it maintains the interrelation of the

two aspects of the universe, spatial and temporal; it also retains the possibility of assigning a definite numerical value to the "interval" between two events, a value which is the same in all systems of measurement.

One point which the new theory emphasizes is that it need not be possible for these numerical measures of the intervals between all events to be fitted together to satisfy the requirements of Euclidean geometry.<sup>1</sup> Thus, when a background or set of co-ordinates has been chosen, and a time scheme has been adopted in an arbitrary manner, the conceptual space intervals may not fit together like ordinary Euclidean elements of length. But, we must remember that the relation of these conceptual space intervals to measured lengths has yet to be discussed (see pp. 126 ff.).

Here it is wise to warn the reader against possible confusions in language. Some would say that the last sentence may be paraphrased thus "Space is non-Euclidean"—implying that for them *Space* is nothing more than an aspect of the metric relations between material phenomena—it is not a background of co-ordinates, it is not the aggregate of all sets of values of three parameters.

This is a logical attitude. If it is taken, the quantities  $g_{\mu\nu}$  are characteristic of and indeed define the relation of the set of co-ordinates chosen to the complex of physical changes.<sup>2</sup> But it is most important to keep it always in mind that the word 'space' is now only a convenient term to cover the metric relations of the physical world.

On the other hand, there are those who would use the word

<sup>1</sup> This is not even true in the case of the special Principle of Relativity; for there the line element is given by

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2$$

and if  $x, y, z, t$  are all real quantities the related geometry is not Euclidean. Only when  $t = \tau\sqrt{-1}$ ,  $\tau$  being real, do we arrive at a quasi-Euclidean geometry. In the actual universe it is generally possible to join two points by a line of zero length, a distinctly non-Euclidean property.

<sup>2</sup> This is presumably what Prof. Eddington means when he says ("Report," p. 22) that "the values of the  $g$ 's express the metrical properties of the space used". Here clearly "space" means co-ordinate system. Later, however, we read, "When he speaks of space, he means the space revealed by measurement, whatever its geometry"—("Space, Time, and Gravitation").

'space' in such a sense that by itself it has no metric properties. It cannot be said to be either Euclidean or non-Euclidean. Given the complex of phenomena we are at liberty to choose out any arbitrarily moving point and say that its course is a straight line defined uniformly—or we may take any curve whatever in the four-dimensional picture and call it straight. We may take any fourfold meshwork as defining a set of parameters. These may be taken as ordinary Cartesian co-ordinates—the picture may be strained, so to speak, until the parametric surfaces are flat. There is a Euclidean geometry in this fourfold background, but that geometry does not correspond with the geometry of measured motion. The Euclidean element of length is not the measured element of length. Over and above the frame of reference we require to know the values of the coefficients  $g$  at each point. From this point of view, to the individual observer who has chosen his frame of reference—once for all—the coefficients  $g$  appear as properties of the physical world relative to his frame of reference. Here we are nearer to the Newtonian position; and provided we recognize the equal validity, that is, the complete relativity, of all space-time partitions of the physical world, there is nothing to hinder.

But the former position is in one respect to be preferred. For it recognizes the physical reality as one, and the mental pictures of it as many, subjective and relative. The latter position requires one to think of quantities such as  $g_{\mu\nu}$  as representing metrical properties of the physical world, although their values only partially depend upon physical phenomena, and indeed are largely arbitrary.

The theory recognizes no physical significance in any equations which are dependent for their validity on the choice of a system of measurement. In that sense it is a theory of the "absolute" physical world, the world absolved from individual views of it. The relativity is characteristic of the view.

## CHAPTER IX.

### FURTHER CRITICISM AND GENERALIZATION. WEYL'S THEORY OF ELECTRICITY.

(The reader will have realized that we are dealing now with an essentially mathematical theory, and it cannot be hoped that this chapter will give any other impression. The main conclusions have been italicized. I have to thank Prof. H. F. Baker for valuable criticism of this chapter.)

#### 97. RECAPITULATION.

IN the preceding chapter the order of the argument has been mainly historical; the purpose has been to give a straightforward account of Einstein's method and results.

Before concluding, it is worth while however to recapitulate and to see the general lines upon which a radical reconstruction of physical theory must proceed. We shall see that even Einstein's generalized relativity is not the last word, and an attempt will be made to indicate the lines along which Weyl has shown that it is possible by a further step to give to electricity as natural a place as that accorded to gravitation in Einstein's theory. Einstein himself has shown that it is possible to generalize the fundamental equations of electrical theory so that they have an invariant form, and are therefore consistent with the Principle of Relativity in the sense in which he uses the term. But in his treatment the connection between the gravitational field and the geometry of the measure system is so intimate, while that between the electrical field and the measure system is so remote, that the fundamental place accorded to electricity in modern thought seems to be denied. In Weyl's theory it is completely restored, and gravitation and electricity rank together as two fundamental properties of matter, forming together the basis of all natural systems of measurement.

#### 98. RECONSTRUCTION.

The universe of phenomena is now conceived to be laid out as a four-dimensional whole. Corresponding to our ordinary

conception of a point of space at a definite instant of time, we have in this four-dimensional whole a marked point or position, which to save confusion it is better to call an event. The history of a particle is recorded as a continuous one-dimensional sequence of such events, or a curve.

For the purpose of distinguishing between events we introduce a set of four variable quantities ( $x_1, x_2, x_3, x_4$ ); to each event corresponds a definite set of values of these. Two events are neighbouring events if the values of these co-ordinates differ only by small quantities.

Our object now is to search out definite relations between these events, to describe by comprehensive formulæ the nature of the sequences which we are able to observe. This is the work of the mathematician, who must accordingly develop his machinery.

#### 99. THE NATURE OF OBSERVATIONS AND OF MEASUREMENT.

It has several times been remarked above that the only precise observations are those of coincidences.<sup>1</sup> The physicist constantly seeks to eliminate from his experiments the application of personal human judgment to the estimation of intervals of time or space. He makes the photographic plate, the micrometer screw, the galvanometer scale do the work for him. Thus the work of observation is reduced to the recording of the fact that a certain event in one chain of events coincides with a certain other event in another chain; or, in other words, of the fact that a certain event is common to two particular world lines. The work of theory consists in discovering the relations that hold between such observations.

But, it may be asked, granted that all operations of measurement may be reduced to observations of coincidences between marks on scales and on the bodies under measurement, between spider lines and optical images, and so on, what is the relation of the intervals of length and time as measured to the co-ordinate numbers which have been introduced merely for the purpose of distinguishing one

<sup>1</sup> See §§ 24, 43, 72.

event from another? What meaning in fact is left for such a term as "a measured interval of length".

If we analyze elementary methods of measuring distance, we come back to the calibration of a scale. For the purpose of doing this, we have to take a certain body, *which is conceived to be permanent in configuration* (though what this really means it is extremely difficult to say); we bring a certain segment of this body between two marks into coincidence with successive segments of the scale, and mark off the lengths with which it coincides each time. These lengths are called equal because they coincide in turn with the same segment of the trial body. The scale itself is conceived also to *be permanent in configuration*. Thus a measuring rod is prepared.

Thus practical measures of length rest on the conception of an ideal "rigid" scale. But all that can be said in definition of the term "rigidity" is, that experiment shows that we can obtain bodies which give approximately consistent measurements. Two segments which have once coincided may be expected to coincide again if brought together. There is no absolute standard by which it may be said of any one body that it always has and will have the same configuration. We can hardly say more than this, that rigidity is the common property of those bodies whose mutual measurements are always and everywhere consistent, but that such bodies do not, in fact, exist.

#### 100. FURTHER DISCUSSION OF THE INTERPRETATION OF THE ECLIPSE OBSERVATIONS.

Now, to make discussion concrete, let us consider in detail the interpretation of the actual observations made by the Eclipse Expedition. These have been said to confirm Einstein's theory. In what sense is this the case? What are the actual steps taken, first in the application of the theory in this concrete instance; second, in the practical measurements?

In § 91, it was said that of all conceivable co-ordinate systems, only one is consistent with the two hypotheses:—

- (i) That the gravitational field does not vary with the time;

- (ii) That the field is symmetrical round the sun as centre  
(By symmetry it was implied, that if the space co-ordinates are treated as Euclidean the gravitational tensor depends only on the distance from the centre.)

On the basis of this set of co-ordinates a picture of the universe was set up. The paths of the rays of light were marked out in this picture on the hypothesis that the equation  $ds = 0$  represented the law of propagation. It was then found that these rays were curved (in the picture) to a greater or less degree according to the distance at which they passed the sun.

The crucial test is: Can this curvature be detected? Let it be emphasized that the curvature is relative to a background of conceptual co-ordinates. The theory proceeds to map out the photographic plate, associating with each point of it values of the Einstein co-ordinates. *The displacements of the images predicted by the theory are stated in terms of changes in the theoretical co-ordinates of the images.* In the actual observations, photographs were taken of a certain group of stars, and the relative positions of the images on the plate were compared with the positions of the images of the same stars photographed at a time when the rays did not pass near the sun. Measurements *experimentally* taken by means of a micrometer measuring-machine showed certain changes in the relative positions *as plotted in terms of readings on the micrometer.* Theory, on the other hand, predicts that there will be changes in the relative positions *as plotted in terms of the conceptual scale of co-ordinates.*

Until we have assured ourselves of the relation between the micrometer measurements and the numbers, which for convenience we may call Einstein's co-ordinates, we cannot know what relation to expect between the changes in the positions of the images in the co-ordinate system and the measured displacements.

As a matter of fact, the measurements of the plates showed a surprising agreement with the predicted displacements in the conceptual picture. What then may we conclude?

We are tempted to say at once—the theory is correct. But there is more than this. It is just conceivable that the

theory might be wrong, but that a further discrepancy between conceptual co-ordinates and measured positions might bring the prediction again into agreement with observation. This seems unlikely in a high degree—so that our conclusion is that there is a considerable probability

- (i) That Einstein's conceptual picture of the phenomena is a consistent one, and
- (ii) That his conceptual measures of distance by co-ordinate numbers are in accord with the experimental measures. To this point we will return shortly.

#### 101. RESUMÉ OF NEWTONIAN CONCEPTIONS.

Let us before going further see how Newton's stock of conceptions stands in relation to what has just been said.

Newton postulates an absolute time and space; but we must do him justice and admit that, his own definitions notwithstanding, he does not anywhere assume the existence of an absolute standard apart from the phenomena themselves. His space and time are a conceptual scheme employed for the purpose of describing the relations between phenomena.

On this conceptual background of time and space co-ordinates he elaborates his picture of the motions of bodies; he is thus able to predict which point of his conceptual space corresponds to a planet at any moment of his conceptual time; he can express the space-co-ordinate numbers of the planet in terms of the time co-ordinate number. When this is done it is found that, except in the long-standing discrepancy in the motion of the perihelion of Mercury, there is precise agreement, within the limits of accuracy to which we can rely upon our measurements, between the positions as predicted by the conceptual co-ordinate numbers and the co-ordinate numbers actually obtained by measurement. That is, astronomical observations confirm the suggestion that *the partitioning of space by means of measurements with a so-called rigid body agrees with the partitioning by means of Newton's conceptual co-ordinates*. This means further that it is confirmed, indirectly it is true, but none the less distinctly,



that the "rigid bodies" involved, scales and micrometer screws, occupy exactly the same volume and shape as described by the conceptual co-ordinates, no matter where they may be placed, whether they are moving or at rest.

Thus the success of Newton's theory, as far as it goes, confirms the belief that there is a conceptual space-time system of measurement in which—

- (i) Newton's laws of motion hold ;
- (ii) Newton's law of gravitation holds ; and
- (iii) The conception of a rigid body is tenable.

In other words, these three hypotheses are mutually consistent in the light of experiment. In itself this conclusion may seem very meagre, and merely a repetition of familiar truths in a different form ; but by stating them in this way we avoid Newton's unphilosophic postulation at the outset of a self-existent absolute space and time ; we grant the *a priori* relativity of all space and time measures ; but at the same time we recognize that the relations existent within the realm of phenomena determine what conceptual space-time system we may *most conveniently* adopt.

## 102. THE RIGID BODY: THE PHOTOGRAPHIC PLATE.

In the analysis that has now been made the co-ordinate system appears merely as a mathematical machinery for comparing certain perceived relations between phenomena.

Among these phenomena are the motions of the planets and the propagation of light. But we must now include as a phenomenon of perception the existence of the approximately rigid body. That is, there is no longer supposed to be any such thing as an ideal or conceptual rigid body. It is simply a matter of experience that in a certain co-ordinate system, chosen for the simplicity which the equations of motion take relative to it, there is a large class of actual bodies which occupy approximately the same space in the picture wherever they may be placed, and at all time. The "ideal rigid body" would be nothing more than a body which did this exactly ; but to speak of such a concept is simply to give a name borrowed from concrete experience to a purely numerical

conception. There is no such thing, therefore, as "measurement by the ideal rigid body". The phrase means nothing more than the comparison of certain perceived phenomena through the mediation of a conceptual co-ordinate system. But the rigid body of experience is essentially a complicated dynamical system. Its property of permanence is of the same kind as the permanence of the solar system as a whole.

Thus we have to recognize that the measuring of the eclipse photographs by the micrometer might quite possibly give results differing from the measures by means of the co-ordinate numbers at which theory arrives.

But just as the system of co-ordinates which is evolved by Einstein as most appropriate to the gravitational field round the sun differs only slightly from that obtained by Newton in its relations to the planetary motions, so it differs only slightly in its relation to the dynamical system which constitutes the photographic plate. It is in fact a matter for observation or experimental verification that a rigid body, in the ordinary sense, a plate of glass, for instance, or the micrometer screw does occupy approximately a constant region as marked out by Einstein's co-ordinates, so that the calibration of the plate by the screw agrees very approximately with that by means of the co-ordinates.

Thus the agreement of the theoretical displacement obtained by Einstein with the actual measurements is as much a confirmation of this property of the rigid body as it is a verification of Einstein's theory of the nature of gravitation. It is a confirmation of the mutual consistency of all the hypotheses, but not strictly a demonstration of any one of them.

### 103. THE MATHEMATICAL MACHINERY FOR DESCRIBING THE ORDER OF A CHAIN OF EVENTS.

The method of the mathematician is first to devise a means of describing in terms of numbers the order or arrangement of a one-dimensional series of events. If it were a discrete series, the work would be merely one of enumeration. But all possible events form for us a continuum, and this requires that to all the events ranged as they are in order along a

world line we shall make correspond the values of a continuously increasing number. This is done by assigning a measure to the separation of two adjacent events. In this connection an infinitesimal calculus is introduced.

Riemann in his discussion of the bases of geometry at once introduces as the measure of separation of two neighbouring events the quadratic form  $ds^2 = g_{\mu\nu} dx_\mu dx_\nu$ ,  $dx_\mu$  being the excess of the co-ordinates of an event P' over those of a neighbouring event P. This is the simplest expression that can be devised which will vanish if P' coincides with P, and which may be positive for all values of the ratios of  $dx_\mu : dx_\nu$ , that is for all events P' in the neighbourhood of P. Here  $g_{\mu\nu}$  are quite arbitrary. The linear form  $f_\mu dx_\mu$  is simpler, but is excluded inasmuch as the changing of sign  $dx_\mu$  would change the sign of the whole expression, and therefore lead to a quantity which would express the separation of two events in some cases as negative.

#### 104. GENERALIZATION.

But Riemann's method is not the most general that can be adopted for the purpose we have in view, namely, the description of the order of events along any arbitrary world line. It is true that taking any world line through a point P, and summing up we may write  $s = \int_P^Q ds$  as the separation of Q from P; since  $ds$  is always positive,  $s$  will increase continuously as Q moves along the world-line without retracing its steps, so that Riemann's method is adequate. But suppose, that having laid down such a set of quantities  $g_{\mu\nu}$ , and thus associated a value of  $ds$  with each interval, we arbitrarily and non-uniformly stretch our measure of the separation of the events along a certain curve. Suppose that, in mathematical language, fixing attention on a certain defined world-line we write

$$d\sigma = \phi ds$$

where  $\phi$  is an arbitrary positive multiplier varying in any manner from point to point along the world-line. Then  $\int d\sigma$  varies always in the same sense as  $\int ds$ ; and therefore for the single purpose of describing the order of events on this world-line  $\int d\sigma$  is as adequate as  $\int ds$ .

Also as long as we are thinking only of one particular world-line we are concerned only with the values of  $\phi$  along that line. We are not bound in considering two world-lines which intersect one another to limit ourselves to the same quantity  $\phi$  at the common event. For definiteness all that is necessary is that each world-line shall itself determine what value of  $\phi$  is to be taken at each point of itself.

Now just as Riemann found a *sufficient* means of prescribing a measure for the interval between adjacent events by means of the quantities  $g_{\mu\nu}$  and the quadratic form, we are able to devise a sufficient means for our new purpose by laying down the rate at which  $\phi$  varies in any direction starting from an arbitrary event. This is done by putting

$$\delta(\log \phi) = \sum \phi_\mu dx_\mu \quad (\mu = 1, 2, 3, 4)$$

where  $\phi_\mu$  is a set of four arbitrary functions.

„ If further we allow that when the co-ordinates are changed  $\phi_\mu$  is also to be so changed that the *linear form*  $\phi_\mu dx_\mu$  remains invariant, then the values of  $\phi$  to which it leads along any world-line are not affected by any change of co-ordinates.

#### 105. EINSTEIN'S INTERVAL BETWEEN TWO EVENTS NO LONGER A DEFINITE QUANTITY.

Suppose then that this linear form is laid down in an arbitrary manner. In general, starting from a given value  $\phi_0$  at P, the value of  $\phi$  at another point Q will depend upon the path by which Q is reached. This will be so unless  $\int \phi_\mu dx_\mu$  taken round any closed curve is zero<sup>1</sup>—that is unless  $\frac{\partial \phi_\mu}{\partial x_\nu} = \frac{\partial \phi_\nu}{\partial x_\mu}$  at all points. But the purpose we have in view does not need that this should be so, inasmuch as we are only concerned with the order of points along one world-line at a time.

The gradients  $\phi_\mu$  rank with the coefficients  $g_{\mu\nu}$ . We may lay down their values arbitrarily for the whole of the field.

<sup>1</sup> For  $\log(\phi_Q/\phi_P) = \int_P^Q \phi_\mu dx_\mu$ . If  $\int_P^Q \phi_\mu dx_\mu$  has the same value for two paths say PAQ, and PBQ; then  $\int \phi_\mu dx_\mu$  for the path PAQBQ is zero; and vice versa.

But when this is done, the multiplier  $\phi$  is not as a rule defined except as to its variation along any given world-line. Thus *the interval*  $d\sigma = \phi ds$  *is not a unique and defined quantity*. Thus with this generalization the argument for a shift in the spectral lines due to a gravitational field does not hold (see § 95).

#### 106. REDUNDANCE IN THE ARBITRARINESS OF THE MODE OF ORDERING.

There is a certain amount of overlapping however in the arbitrariness of the  $g_{\mu\nu}$  field and the  $\phi$  field.

We do not alter the value of the separation  $d\sigma$  for every element of any given curve if we multiply the value of  $\phi$  by an arbitrary function of position  $\psi$ , and at the same time divide each of the quantities  $g_{\mu\nu}$  by  $\psi^2$ .

Thus the ordering of events by this method is one in which the measure relations remain unaltered if we put

$$g_{\mu\nu}' = g_{\mu\nu}/\psi^2$$

$$\phi_\mu' = \phi_\mu + \frac{\partial}{\partial x_\mu} (\log \psi).$$
<sup>1</sup>

Further we do not change the value of each element of separation if we replace the co-ordinates  $x_\mu$  by any other set of co-ordinates  $x'_\mu$  provided that at the same time the coefficients  $\phi_\mu$  and  $g_{\mu\nu}$  are replaced by suitable other quantities  $\phi'_\mu$  and  $g'_{\mu\nu}$  such that the two differential forms  $\phi_\mu dx_\mu$  and  $g_{\mu\nu} dx_\mu dx_\nu$  are both invariant. In the language of the ordinary tensor theory employed by Einstein, this means that  $\phi_\mu$  is to be a covariant vector and  $g_{\mu\nu}$  the ordinary covariant tensor of second rank.

*We may say that  $\phi_\mu$  and  $g_{\mu\nu}$  define a measure-system, but a given measure-system leaves  $\phi_\mu$  and  $g_{\mu\nu}$  largely arbitrary.*

<sup>1</sup> Putting

we have

or

i.e.

$$\phi' = \psi\phi,$$

$$\delta\phi'/\phi' = \delta\psi/\psi + \delta\phi/\phi$$

$$\phi'_\mu \delta x_\mu = \frac{\partial}{\partial x_\mu} (\log \psi) \delta x_\mu + \phi_\mu \delta x_\mu.$$

$$\phi'_\mu = \phi_\mu + \frac{\partial}{\partial x_\mu} (\log \psi).$$

# 107. NATURE DOES NOT MAKE QUANTITATIVE EVALUATIONS.

Thus much for the mere machinery of recording the order of events. Next as to the nature of the relations which describe natural phenomena.

The observed relations, as has been said, are merely of coincidence. But they are always mathematically recorded by means of numerical quantities introduced for the purpose. *Nature itself cannot recognize or prescribe any particular method of quantitative evaluation*; that is only part of our intellectual machinery.

If, therefore, we write down any mathematical equation which expresses observed relations it must be of such a nature that its validity is independent of the arbitrary choices of the quantities used for the purpose of mathematical expression.

It should be true, therefore, that any order relations will be maintained true if we make any arbitrary change in the method by which we assign to any pair of events a numerical measure of the interval between them.

## 108. INTRINSIC QUALITIES OF A MEASURE-SYSTEM. DIRECT ROUTES. INVARIANTS.

If we adopt an arbitrary measure-system, that is, if we assign certain numerical magnitudes to the intervals of each world-line, as in §§ 104-6 we have a complete relativity of co-ordinates within such a measure-system, as well as an arbitrariness in what we may call the *gauge-system*  $\phi_\mu$ .

But *each measure-system has intrinsic characteristics independent of the choice of co-ordinate-system or gauge-system*. As an illustration of this, suppose that we have assigned an arbitrary field of values to  $g_{\mu\nu}$  and  $\phi_\mu$ , and adopted an arbitrary system of co-ordinates for the mapping out of our four-dimensional world. Let A, B be any two events, and  $\phi_0$  the value of  $\phi$  at A. Then for any given path between A and B,  $\int d\sigma$  is defined. It follows that the *direct routes* from event to event are

thereby laid down, a **direct route** being defined as one for which  $\int d\sigma$  has a stationary value. Thus each measure-system has its characteristic system of direct routes or geodesics, the determination of which depends only upon the two invariants  $\phi_\mu dx_\mu$  and  $g_{\mu\nu} dx_\mu dx_\nu$ .

Each measure-system may also have associated with it certain quantities whose values are unaltered under co-ordinate changes and guage-changes which leave any given interval unaltered. For instance the integral

$$\iiint f_{ik} f^{ik} \sqrt{g} dx_1 dx_2 dx_3 dx_4$$

where  $f_{ik} = \frac{\partial \phi_i}{\partial x_k} - \frac{\partial \phi_k}{\partial x_i}$ , and  $f^{ik} = g^{ia} g^{kb} f_{ab}$

taken over any given region in the four-dimensional world is such an invariant. Another quantity so invariant is  $\int \phi_\mu dx_\mu$  taken round any closed curve; for, on the understanding that  $\phi_\mu$  is a covariant vector  $\phi_\mu dx_\mu$  is invariant for co-ordinate changes, and for a change of guage as above (§ 106), we have

$$\begin{aligned} \int \phi'_\mu dx_\mu &= \int \phi_\mu dx_\mu + \int \frac{\partial \phi}{\partial x_\mu} (\log \psi) dx_\mu \\ &= \int \phi_\mu dx_\mu \end{aligned}$$

since the total increment in  $\log \psi$  is zero on describing a closed circuit.

#### 109. PHYSICAL LAWS AND SPECIAL MEASURE-SYSTEMS

Einstein's theory and its extension at the hands of Weyl seek now to establish a correspondence between these intrinsic properties of the measure-system and the universe of material phenomena.

For instance, part of Einstein's hypothesis for the explanation of planetary motion is that *the free motion of a particle is a direct route or geodesic of the measure-system*. We may put this hypothesis in the form of a question—*can we, out of all possible measure-systems, select one in which the direct routes are the actual paths of free particles?*

Further, Einstein seeks to set up a correspondence between

the presence of matter, the gravitational influence of matter and the intrinsic properties of the measure-system; with this in view he selects the tensor,  $G_{\mu\nu}$ , propounding the hypothesis that the vanishing of this is to correspond with the absence of matter.

Thus we have another question; *is it possible to select from all possible measure-systems one which is such that the tensor,  $G_{\mu\nu}$ , vanishes at all regions in the four-dimensional world which are not occupied by matter?*

To both of these questions the provisional answer is "yes"; and the experiments have confirmed the answer by showing that its consequences in the region, open to test, are actually observed. Thus *the law of gravitation becomes the condition which singles out a measure-system*. We have seen in the last chapter how the law  $G_{\mu\nu} = 0$ , with the assumption of symmetry round a centre, did actually lead to a definite measure-system in which the path of Mercury is verified in practice to be a geodesic.

#### 110. NATURAL MEASURE-SYSTEMS.

What does this mean for other measure-systems in which the path of Mercury is not a direct route? Mathematically they are equally capable of describing the relative positions and orders of the four-dimensional world. But in the picture which they would give of it, there would be a confusion between the natural geodesics or freepaths of particles, and the direct routes of the measure-system; between the regions occupied by matter and the regions of curvature of the measure-system. The confusion would be of the same type as that which in the old Newtonian theory would exist between actual forces producing accelerations and apparent forces arising if the motion of the frame of reference was unknown. Thus it appears that among all possible measure-systems there are some in which there is a specially simple form for the order relations inherent in the physical world. These might be called *natural measure-systems*.

Prof. Eddington states the conclusion with regard to natural measure-systems differently. What has been called



here a measure-system, he calls a "species of space-time". He asks if every kind of space-time can occur in nature. To this he replies, "No! Einstein's law is a law which determines those kinds of space-time which can occur in nature". This suggests that material phenomena are incapable of description on the background of any kind of space-time not satisfying that law. But if the view expressed above is correct, this is not so.

There is of course no such thing as a *measured space-time*. We may speak of measuring space in the ordinary experimental sense; but the process of measuring is itself a sorting out of certain relations in the complex of phenomena. In the four-dimensional world in which all phenomena are laid out as a whole, any experiment of measuring space is laid out as part of the picture, and the observations are simply of the fact that certain world lines intersect (see § 99).

We must interpret Prof. Eddington's assertion that "only certain kinds of space-time occur in nature" in the sense that only certain kinds show that close correspondence with nature which enables us to identify matter with aspects of curvature. But we must remember that, in making such an identification, we are simply resuming in picturesque language the fact that experiment confirms Einstein's law. Curvature of a four-dimensional manifold is a mere arithmetical conception, it is form without content. The process of identification is simply a recognition that the order relations of the physical world are of that mathematical type which we find in the algebraic theory which, by analogy with the early ideas of curvature, we call the theory of curvature in four-dimension or five-dimension space.

### III. LEAST ACTION.

It has been remarked in chapter vii., p. 110, that Einstein's law of gravitation follows from the principle of least action, if we take as the action the expression  $\iiint G \sqrt{g} \, dx_1 \, dx_2 \, dx_3 \, dx_4$ ,  $G$  being the scalar invariant of curvature, which, we may remind ourselves, is a function of  $g_{\mu\nu}$  and their first and second

differential coefficients. In applying the principle of least action the method is to calculate the variation of this expression for arbitrary changes in  $g_{\mu\nu}$ .

Writing  $\delta A = \iiint G_{\mu\nu} \delta g^{\mu\nu}$ , it is inferred that if  $\delta A = 0$  for an arbitrary small change in the tensor  $g^{\mu\nu}$ , then we must have  $G_{\mu\nu} = 0$ . This stands as an equation, limiting the arbitrariness otherwise allowable in the tensor  $g_{\mu\nu}$ .

The principle of least action in its physical aspect would then say that the happenings in nature are such as, when interpreted in terms of a gravitational field  $g_{\mu\nu}$ , cause that field to be such that this function of its whole condition, the action has a least (or rather stationary) value.

On the geometrical side, however, if we think of  $g_{\mu\nu}$  with a set of co-ordinates as constituting a measure-system, the law simply stands as a principle of selection among the measure-systems. We arrive at the conclusion that, *in order to obtain the simplest and most complete correspondence between geometry and natural law, we have to choose our measure-system in such a way that the total curvature,  $\iiint G \sqrt{g} dx$  has a least value,*

Here of course we are not considering the further generalization of this chapter.

## 112. WEYL'S THEORY OF ELECTRICITY.

In Einstein's theory of gravitation as described in the last chapter, a certain expression  $G_{\mu\nu}$  depending on  $g_{\mu\nu}$  and its differential coefficients was described. The vanishing of  $G_{\mu\nu}$  at a point, indicated the absence of matter at that point. The coefficients  $g_{\mu\nu}$  may be taken as characteristic of the gravitational field, and  $G_{\mu\nu}$  as measuring the nature of the matter at any point.

We have now contemplated four other quantities  $\phi_\mu$ . In their origin, like  $g_{\mu\nu}$ , they arose out of the attempt to express order relations in a quantitative form. But the possibility of establishing a close correspondence between the properties of matter and those of the measure-system enabled us to give to  $g_{\mu\nu}$  a physical significance. In the same way any

correspondence that may be set up and experimentally confirmed between some property of the measure system involving  $\phi_\mu$  and some property of matter will enable us to give a physical significance to  $\phi_\mu$ .

Now it was mentioned above that  $\int \phi_\mu dx_\mu$  taken round any closed curve is a quantity fixed within a given measure-system. There may for instance be regions which are such that this quantity vanishes for any closed curve drawn within the region. This will be a property of such a region whatever changes are made in the co-ordinate system or in  $\phi_\mu$ , subject only to the measures of the intervals being unvaried.

But if the  $g_{\mu\nu}$  and  $\phi_\mu$  are varied in an *arbitrary* manner, clearly  $\int \phi_\mu dx_\mu$  will not be an invariant.

Now in Weyl's theory *any region which is such that  $\int \phi_\mu dx_\mu$  vanishes for all curves drawn within it corresponds to one which is free from electric and magnetic force.* Effectively this means that the four quantities are identified with the electro-magnetic potentials,<sup>1</sup> or *conversely that in a field which is free from electric and magnetic force  $\phi$  has a unique value at end point, and therefore the interval  $d\sigma$  has a definite value.*

Here as in the gravitational case we are recognizing simply a community of form, which makes it possible for us to identify intrinsic qualities of a measure-system with qualities of the physical universe. Their values will be relative to the co-ordinate system adopted.

<sup>1</sup> Thus: if we put  $f_{ik} = \frac{\partial \phi_i}{\partial x_k} - \frac{\partial \phi_k}{\partial x_i}$ , the necessary and sufficient conditions that  $\int \phi_\mu dx_\mu = 0$ , for any curve within a region are that  $f_{ik} = 0$  for all values of  $i$  and  $k$ , at all points of the region.

Putting  $\phi_\mu = (F, G, H, \text{ and } \Phi)$  we have

$$\begin{aligned} f_{14} &= \frac{\partial F}{\partial x_4} - \frac{\partial \Phi}{\partial x_1}, & f_{23} &= \frac{\partial H}{\partial x_3} - \frac{\partial G}{\partial x_2} \\ f_{24} &= \frac{\partial G}{\partial x_4} - \frac{\partial \Phi}{\partial x_2}, & f_{31} &= \frac{\partial F}{\partial x_3} - \frac{\partial H}{\partial x_1} \\ f_{34} &= \frac{\partial H}{\partial x_4} - \frac{\partial \Phi}{\partial x_3}, & f_{12} &= \frac{\partial G}{\partial x_2} - \frac{\partial F}{\partial x_1} \end{aligned}$$

which enable us to identify  $F, G, H$ , with the vector potential and  $\Phi$  as the scalar potential,  $(f_{14}, f_{24}, f_{34})$  being the electric intensity and  $(f_{23}, f_{31}, f_{12})$  the magnetic intensity.

Further, in the regions in which  $\int \phi_\mu dx_\mu$  round a closed curve does not vanish, it is possible out of the functions  $\phi_\mu$  to construct a farther set of quantities  $S_\mu$ , the relation of which to  $\phi_\mu$  is identical with the relation of the stream vector  $\rho$  ( $u_x, u_y, u_z, c$ ) to the electro-magnetic potentials (see p. 84). The relation between  $S_\mu$  and  $\phi_\mu$  is preserved in any arbitrary change of the measure system. The tensor  $S_\mu$  represents an intrinsic quality of the measure-system.

Weyl's hypothesis is that *it is possible so to choose the measure system that the regions in which  $S_\mu = 0$  correspond exactly to the regions in which there is no electricity*. This is the complete analogue of Einstein's hypothesis that  $G_{\mu\nu} = 0$  corresponds to the absence of matter.

The justification for Weyl's hypothesis is simply the community of form between the relations  $S_\mu = 0$  and Maxwell's equations. They are indeed identical. So far there is no experimental test proposed of the accuracy of the hypothesis. But there is much to recommend it. The main points in its favour are these:—

- (i) The analysis proposed is forced upon us by the incompleteness of Riemann's point of view.
- (ii) The attempt to carry out the generalization makes a natural place for a tensor  $\phi_\mu$  corresponding to the electro-magnetic potentials in addition to the tensor  $g_{\mu\nu}$  which we have seen can represent the gravitational field.

Thus electricity and gravitation are placed side by side as fundamental properties of our perceived universe, common to all kinds of matter.

It will to some extent be a matter of the predilection of the student of these matters, whether he will now consider that the physical laws of the universe are reduced to geometrical laws or the reverse. It is clear that the course of nature is more closely related to some systems of measurement than to others—just as it was when all our exact laws were comprised in Newtonian dynamics. On the other hand, the fact that we are able to make such an excellent guess at the

form of the laws of nature by mere geometrical considerations, leaves us with an uneasy feeling that we are looking at nature through coloured glasses, and are simply seeing those aspects of phenomena which are the easiest for us to appreciate. This however is the usual course of science, and it is something to feel that this most comprehensive of physical theories itself suggests that when our power of vision is greater we shall discover whole new realms of phenomena.

### 113. THE CONSERVATION OF ELECTRIC CHARGE.

It has been said that in Weyl's theory a certain tensor  $S_\mu$  is identified with the stream vector of electricity. Now we find that when the variables  $\phi_\mu, g_{\mu\nu}$  and the co-ordinates are altered in any manner consistent with leaving the interval measures unchanged, the tensor  $S_\mu$  satisfies an identical invariant relation. This is completely analogous to the four identities connecting the quantities  $G_{\mu\nu}$  when the element  $g_{rs} dx_r dx_s$  is left invariant (*cf.* § 89). These four relations, it was said above, are the relations which embody the principles of conservation of energy and momentum, energy and momentum being special aspects of the tensor  $G_{\mu\nu}$  which emerge when we resolve our four-dimensional world into the special aspects space and time. So here, when  $S_\mu$  is interpreted in the space-time aspects of electric charge in motion, *the identical relation which it satisfies as a consequence of relativity stands as expressing the principle of the conservation of electric charge.*

### 114. WEYL'S THEORY AND THE ATOMIC NATURE OF ELECTRICITY.

At this point we become conscious of the limits of Weyl's theory as so far expounded.

The hypothesis is that it is possible to choose certain measure-systems so that certain intrinsic properties of the system correspond to the world lines of the electric charges.

In the light of the electron theory, the function of the equations of the electric field is to co-ordinate the motions of electrons, to predict the motion of an electron in a given field. If we think of electricity as distributed in nuclei of extremely small dimensions, we shall have the equations

$$S_{\mu} = 0$$

as characteristic of the field external to the electrons. These being identical in form with Maxwell's equations will lead to the usual expressions for the field due to electrons having prescribed motions.

But it has already been pointed out (p. 24) that those equations are insufficient to determine the motion of any electron without some further information as to the structure of the electron (e.g. Lorentz' contracting electron) or some equivalent hypothesis. We have seen that the Special Principle of Relativity was able to fill the gap for the slowly accelerated electron.

If we seek to supplement Weyl's equation by some further hypothesis, we have this elementary treatment to guide us, but we have also the method adopted by Einstein with such success for the motion of the material particle. We recall that having by means of the equations  $G_{\mu\nu} = 0$  determined the field of a certain specified distribution of matter, the path of the free particle was taken to be a geodesic of this field.

This would be a possible method of attack in the electrical case. Having obtained the electric field of a given distribution of matter and electricity (we cannot now separate the two) by means of the equations  $S_{\mu} = 0$ , we could make the provisional hypothesis that in this field the world line of a free electron is a direct route, that is a path for which  $\int d\sigma$  is stationary.

This would make the problem of determining its motion a determinate one, would involve no special analysis of the constitution of the electron, and it would be consistent with the results of the special principle which have been verified experimentally. But this is speculation, and the suggestion has not yet been explored.

### 115. SUMMARY OF MATHEMATICAL SPECIFICATION OF WEYL'S THEORY.

$$f_{ik} = \frac{\partial \phi_i}{\partial x^k} - \frac{\partial \phi_k}{\partial x^i}.$$

$$f_{ik} = g^{i\lambda} g^{k\mu} f_{\mu\lambda}.$$

$$\mathfrak{f}_{ik} = \sqrt{g} f_{ik}.$$

Definition of stream of electricity

$$S_i = - \frac{\partial \mathfrak{f}^{ki}}{\partial x^k}.$$

Compare with this Minkowski's equations, p. 84.

These give  $\frac{\partial S_i}{\partial x^i} = 0$

which is the equation of conservation of electricity.

If there is no gravitation  $g_{rs} = 0$  if  $r \neq s$ , also  $g_{11} = g_{22} = g_{33} = -1$ , and  $g_{44} = +1$ . The equations then reduce exactly to Maxwell equations.

The expression for the action in a region not occupied by charge is

$$\int \frac{1}{2} f_{ik} \mathfrak{f}^{ik} d\tau$$

and this integral is proved to be an absolute invariant for any change in co-ordinates, or in the potentials  $\phi_\mu$ , subject to the intervals being unvaried.

### 116. COMPARATIVE SURVEY.

It may be worth while to set down here a comparison of the conclusions from Newton's theory of dynamics, Einstein's restricted Principle of Relativity, Einstein's generalized theory, and finally of Weyl's theory.

I. NEWTON.—In the region of dynamics, of all possible co-ordinate systems there is a limited group for which the laws of motion have a particularly simple form. In any system of the group the path of a free particle is a straight line described uniformly.

II. EINSTEIN, 1905—*Restricted Principle of Relativity*.—In the region of dynamics and electro-dynamics, of all possible co-ordinate systems there is a limited group for which the laws are of precisely the same form. In any system of this group the path of a free particle is a straight line described uniformly, and light travels with constant velocity.

III. EINSTEIN, 1915—*General Principle of Relativity*.—In the region of gravitational phenomena, there is no restriction on the co-ordinate system, but the measure-system is limited by the hypothesis (i) that the world-line of a free particle is a direct line of the measure-system, (ii) that the curvature  $G_{\mu\nu}$

may be identified with the presence of matter. The interval between two events is a fixed quantity within the appropriate measure-system.

IV. WEYL'S THEORY.—In electro-dynamic as in gravitational phenomena, there is no restriction on the co-ordinate system, and the measure of the interval between two events is not a definite quantity determined by the gravitation field. But of all possible measure-systems, there is one, or a limited group, in which the singular regions of the system are in exact correspondence with the regions in which matter or electricity are present.

We conclude then that Einstein has satisfied the demand that nature itself shall not show any preference for any particular system of variables or co-ordinates by which we shall distinguish between events. Further, Weyl has made it clear that the laws of nature *do not supply* us with an absolute criterion of equality of intervals of space and time, except that of coincidence; in fact the idea of a definite measurable interval between two events has to some extent broken down. But we are not left with the conclusion that we may adopt an absolutely arbitrary system of measurement. There is necessarily one measure-system or a limited group, *within which there is complete relativity of co-ordinates*, in which there is a particularly simple correspondence between the geometry of the system and nature; or in other words, for which the mathematical relations of nature take the simplest possible form.

In the end we must admit therefore that the recognition of order in the sequences of events around us arises from an adjustment of mental machinery to the events. Just as we only recognize a star as a sharp point of light when the eye focusses the light on the retina, so we recognize a distinct order in the universe when we focus our measuring system properly. The eye has an infinity of ways of focussing itself so that the star produces a blurred image on the retina. So the mind may form many images of nature which are in a sense blurred, in which a distinct order is not perceptible. But the fact that the eye can produce a sharp image is enough



both to determine our conception of the star, and to define what we mean by the proper focus for the star. So the problem of science is mainly to discover that mental focus in which nature gives us clear impressions of order. Einstein has shown that this may be done in a way far more comprehensive than had been thought possible. He has shown us that the required mental focus is to be attained without any introduction of metaphysical notions such as those of an absolute space and time, or even of such a remnant of those notions as that light shall have a definite velocity the same at all points.

## INDEX.

- ABERRATION**, 16.  
**Abraham**, 38, 70, 81.  
**Absolute**, time, space, velocity, 8.  
**Action**, principle of, 77, 83, 108, 110, 138.  
**Addition of velocities**, Einstein, 39.  
     — — — Newton, 11.  
**Æther**, concealment of, 48.  
     — mechanics of, 89.  
     — momentum in, 81.  
     — penetrates matter, Fresnel, 16.  
**Arago**, 13.  
**Atomic nature of electricity**, 52, 142.  
  
**BESTELMEYER**, A., 66.  
**β-rays**, 64.  
**Brace**, D. B., *see* Rayleigh, 56.  
**Bucherer**, A., 66.  
  
**CATHODE-RAYS**, 38, 68.  
**Christoffel**, 100.  
**Clifford**, W. K., 100.  
**Coincidence**, in experimental observation, 29, 50, 93, 126.  
**Conservation**, of electricity, 48, 142.  
     — of energy and momentum, 12, 79, 111.  
     — of mass, 48.  
**Contracting electron**, 25, 56, 70.  
**Contraction hypothesis**, 22, 23, 44.  
     — Larmor, 24.  
     — Lorentz, 24, 55.  
**Convection coefficient**, 41, 43.  
**Co-ordinate systems**, and propagation of light, 95.  
     — — relativity of, 93.  
**Couple on moving condenser**, 59.  
**Curl of a vector**, 1.  
**Curvature**, in four dimensions, 103.  
     — invariant of, 108-9.  
     — specific, 102.  
  
**DEFLECTION** of light, 119.  
**De Sitter**, 86.  
**Direct rays**, 135.  
**Dispersive media**, convection coefficient in, 43.  
**Doppler effect**, 42.  
**Dynamics**, Newtonian, 8, 10, 39, 78, 83.  
  
**ECLIPSE Expedition**, 121, 127, 131.  
**Economy of thought**, 90.  
**Eddington**, 50, 112, 121, 137.  
**Electric intensity**, relativity of, 46.  
**Electric intensity**, transformation of, 48.  
**Electricity**, 46.  
     — atomic nature of, 52, 142.  
     — conservation of, 48, 142.  
     — Weyl's theory of, 139.  
**Electro-magnetic field**, equations, 46, 84.  
     — inertia, 25, 70.  
**Electron**, apparent mass of, 68, 70.  
     — finite size of, 53.  
     — path of, 143.  
     — theory of matter, 24, 52.  
**Energy**, and mass, 78.  
     — and momentum, 79.  
     — conservation of, 111.  
**Eötvos**, 87.  
**Euclidean geometry**, 102.  
  
**FITZGERALD**, contraction hypothesis, 22, 31, 57.  
**Fizeau experiment**, 14, 16, 41.  
**Force**, and work, 78.  
     — generalized, 78.  
     — mechanical, 45.  
     — on moving charged bodies, 59.  
     — relativity of, 96.  
**Four-dimensional geometry**, 103.  
**Fresnel**, and Arago, 13.  
     — convection coefficient, 14, 41.  
  
**GALILEO**, 78.  
**GAUSS**, 109.  
**Geodesic**, 107, 143.  
**Geometry**, bases of, Riemann, 100.  
     — Euclidean, 102.  
     — four-dimensional, 103.  
     — non-Euclidean, 123.  
**Gravitation**, and electrical theory of matter, 27, 86.  
     — and light, 89.  
     — as a means of communication, 38.  
     — Einstein's Law, 104, 105.  
     — real and simulated, 97.  
**Gravitational field** round the sun, 113.  
     — mass, 87.  
**Gauge-system**, 135.  
  
**HAMILTON's principle**, 108, 110.  
**Hertz**, 17.  
**Hupka**, 68, 70.  
**Huyghens**, 78.  
**Hypothesis of constant light-velocity**, 28.

- INERTIA-mass, 87.  
Interval between events, 121, 133.  
Invariant, action, 83, 112.  
— charge, 12.  
— element of integration, 109.  
— of curvature, 108, 139.
- KAUFMANN's experiments, 25, 64.
- LARMOR, contraction hypothesis, 24.  
— point electrons, 53.  
— Rayleigh-Brace experiment, 57.  
— space-time transformations, 35.  
Length, dependent on velocity, 31.  
Levi-Civita, 100.  
Light, constant velocity, 33.  
— deflection of, 119.  
— gravitational effect on, 89.
- MAGNETIC intensity, relativity of, 46.  
Maps, 91.  
Mass and energy, 80.  
— apparent, of electrons, 67.  
Matter, gravitational field in absence of, 105.  
— modified by motion through the æther, 14, 17.  
— tensor, 111.  
Maxwell, 17, 22, 87.  
Maxwell's equations in Weyl's theory, 141.  
Measure-systems, 134.  
— — correspondence with nature, 136.  
Measuring-rod, 127.  
Mechanical theory derived from electrical, 61.  
Mechanics, 78.  
Mercury, motion of perihelion, 113, 117.  
Michelson, A. A., repetition of Fizeau's experiment, 6, 42, 44.  
Michelson-Morley experiment, 17, 19, 21, 23, 37, 54.  
Minkowski, 72, 83, 84, 94.  
Momentum and energy, 79.  
— conservation of, 12, 111.  
— of æther, 81.  
Morley, E. W., *see* Michelson, A. A.  
— and Miller, D. B., 18, 22, 23.  
Motion, analysis of concept, 10.  
— of charged particle, 62.  
— of free particle and propagation of light, 96.  
— of material particle, 80.
- NEUMANN-Schaefer, 68, 71.  
Newtonian dynamics, 8, 10, 38, 129, 144.  
Newtonian potential, 99.  
Noble, *see* Trouton, 39, 58.  
Non-Euclidean geometry, 123.  
Nordstrom, 86.
- OBJECTIONS to the Principle of Relativity, 38, 41.  
Observations, nature of, 93, 126.
- PATH of free particle has stationary length, 96, 106.  
Perihelion of Mercury, 113.  
Planet, law of motion of, 115.  
Poincaré, 86.  
Potential, Newtonian, 99.  
Poynting, flux of energy, 81.  
Propagation of light, 95-6.
- QUADRATIC Form, 94.
- RANKINE, A. O., *see* Trouton, F. T., 57.  
Rayleigh, Lord, and Brace, D. B., 56.  
— — — on double refraction in moving bodies, 25.  
Relative velocity, 40.  
Riemann, 100, 132.  
Ricci, 100.  
Rigid body, 130.
- SCHAEFER-Neumann, 68, 71.  
Simultaneity, 29.  
Soddy, 41.  
Space and time, absolute, 8.  
— — — unified, 75.  
Space-time transformations, Lorentz, 34, 73.  
— — — Newtonian, 10.  
Spectral-shift, suggested, 121.  
Stokes, Sir G. G., on aberration, 16.  
Sun, gravitational field of, 113.
- TENSOR, defining kind or space-time, 97.  
— of matter or energy, 111.  
Thomson, J. J., 70.  
Transformations, Lorentz, 34, 73.  
— Newton, 10.  
— of charge, 48.  
— of electrical intensity, 47.  
Trouton, and Noble, 39, 58, 82.  
— and Rankine, 57.
- VECTORS, definitions and notation, 1.  
— in four dimensions, 72.  
Velocity, absolute, 8.  
— addition of, Einstein, 39.  
— — — Newton, 11.  
— of light, constant, 33.  
— — — critical, 40.  
— — — in moving matter, 13.  
Velocity, relative, 40.  
Verification of Einstein's theory, 112.
- WEYL, 125, 139, 142, 145.  
Wolz, 66.  
World-line, length of, 106.  
Wright, J. E., 100.
- ZERMAN, and Fizeau's experiment, 44.  
— effect, 68.  
— Eötvös' experiment, 88.





